



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 120900**

**TO: Jeffrey Parkin**  
**Location: rem/3d39/3c18**  
**Art Unit: 1648**  
**Monday, May 03, 2004**

**Case Serial Number: 09/775964**

**From: Edward Hart**  
**Location: Biotech-Chem Library**  
**REM-1A55**  
**Phone: 571-272-2512**

**edward.hart@uspto.gov**

### **Search Notes**

Examiner Parkin,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

**THIS PAGE BLANK (USPTO)**

STIC-Biotech/ChemLib

120900

From: Parkin, Jeffrey  
Sent: Sunday, May 02, 2004 2:02 PM  
To: STIC-Biotech/ChemLib  
Subject: U.S. Serial No. 09/775,964

Please search the following SEQ ID NOS. v. all relevant databases (including interference): SEQ ID NOS.: 5 and 26.

Place results on both paper and disk.

Thanks!

JSP  
Au 1648  
REM 3D39  
2-0908

RECEIVED  
MAY - 3 2004  
STIC/CHEN, Division  
(STIC)

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 5/3/04  
Date Completed: 5/3/04  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:

NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 13:05:13 ; Search time 72.4946 Seconds  
(without alignments)

1781.157 Million cell updates/sec

Title: US-09-775-964-5

Perfect score: 2398

Sequence: 1 PDLRFTNIGPDMRTVWAP.....LVTLPNHLGPEILDVPST 457

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*

1: Geneseqp1980s.\*

2: Geneseqp1980s.\*

3: Geneseqp2000s.\*

4: Geneseqp2001s.\*

5: Geneseqp2002s.\*

6: Geneseqp2003as.\*

7: Geneseqp2003bs.\*

8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2398	100.0	457	2 AAW33340	Aaw33340 Human fib
2	2257	94.1	432	2 AAR40160	Aar40160 Human FN/
3	2257	94.1	432	2 AAW33339	Aaw33339 Human fib
4	2257	94.1	432	2 Aay05457	Aay05457 Fibronect
5	2257	94.1	432	6 ABP72002	Abp72002 Human fib
6	2257	94.1	455	6 ABP72004	Abp72004 Human fib
7	2245	93.6	432	2 AAR40162	Aar40162 Human FN/
8	1536.5	64.1	574	2 AAR08044	Aar08044 Fibronect
9	1536.5	64.1	574	2 AAR60351	Aar60351 Inhibitor
10	1536.5	64.1	574	2 AAW33349	Aaw33349 Oligopept
11	1536.5	64.1	574	2 AAW13572	Aaw13572 Escherich
12	1536.5	64.1	574	2 AAY05453	Aay05453 Fibronect
13	1536.5	64.1	574	2 AAW97357	Aaw97357 Protein s
14	1536.5	64.1	574	6 ABP59436	Abp59436 Human fib
15	1536.5	64.1	574	7 ADD49015	Add49015 Fibronect
16	1527	63.7	489	2 AAW33343	Aaw33343 Protein u
17	1523	63.5	573	2 AAR60350	Aar60350 Chimeric
18	1501	62.6	2265	4 AAM38647	Aam38647 Human pol
19	1485.5	61.9	302	2 AAR60355	Aar60355 Chimeric
20	1485.5	61.9	302	2 AAR98816	Aar98816 Cell adhe
21	1485.5	61.9	302	2 AAW33352	Aaw33352 Oligopept
22	1485.5	61.9	302	6 ABP59437	Abp59437 Human fib
23	1485.5	61.9	302	7 ADD49016	Add49016 Fibronect
24	1474	61.5	2320	6 AAE37107	Aae37107 Human fib
25	1474	61.5	2320	6 ABR42588	Abr42588 Human fib

26	1474	61.5	2320	6 ABO01288	Abo01288 Human pro
27	1474	61.5	2324	2 AAR92778	Aar92778 Human fib
28	1474	61.5	2324	5 AAU74674	Aau74674 Human fib
29	1474	61.5	2324	5 AAE23651	Aae23651 Human pro
30	1474	61.5	2327	1 AAP70373	Aap70373 Human fib
31	1474	61.5	2327	2 AAR15468	Aar15468 Human fib
32	1474	61.5	2328	4 AAG68182	Aag68182 Fibronect
33	1474	61.5	2328	6 ABU07486	Abu07486 Protein d
34	1474	61.5	2328	6 ABR41106	Abr41106 Human fib
35	1474	61.5	2328	6 ABR92078	Abr92078 Human cer
36	1474	61.5	2328	7 ADB70378	Adb70378 Fibronect
37	1474	61.5	2328	7 ADB98726	Adb98726 Human fib
38	1474	61.5	2328	7 ADE82522	Ade82522 Human pro
39	1474	61.5	2355	6 ABR58335	Abr58335 NM_00202
40	1474	61.5	2386	2 AAW63171	Aaw63171 Amino aci
41	1474	61.5	2386	5 AAO17353	Aao17353 Human fib
42	1474	61.5	2386	6 ABR81866	Abr81866 Human fib
43	1474	61.5	2386	7 ADD18770	Add18770 Human dis
44	1474	61.5	2386	7 ADE63324	Ade63324 Human pro
45	1474	61.5	2446	2 AAR60021	Aar60021 Fibrinoge

#### ALIGNMENTS

RESULT 1

AAW33340

ID AAW33340 standard; protein; 457 AA.

AC AAW33340;

XX

DT 23-FEB-1998 (first entry)

XX

DE Human fibronectin fragment containing CS-I binding domain.

XX

KW CS-I binding domain; human fibronectin; target cell; transfection;

KW retroviral vector; gene therapy; cancer; viral disease;

KW acquired immunodeficiency syndrome; AIDS.

XX

OS Homo sapiens.

XX

PN WO9718318-A1.

XX

PD 22-MAY-1997.

XX

PF 07-NOV-1996; 96WO-JP003254.

XX

PR 13-NOV-1995; 95JP-00294382.

XX

PR 08-MAR-1996; 96JP-00051847.

XX

(TAKI ) TAKARA SHUZO CO LTD.

XX

Asada K, Uemori T, Ueno T, Koyama N, Hashino K, Kato I;

XX

WPI; 1997-289294/26.

DR N-PSDB; AAT93970.

XX

Method for increasing efficacy of gene transfer to target cell using

retrovirus - by infection of the target cell in the presence of a

substance which binds to the virus and a substance which binds to the

target cell.

XX

Claim 42; Page 97-99; 194pp; Japanese.

XX

The present sequence is a human fibronectin fragment containing a CS-I binding domain, which was used in the development of a novel method for increasing the efficiency of gene introduction into a target cell using a retroviral vector. The method comprises carrying out viral infection of the target cell in the presence of a retrovirus and target cell binding substance or substances. The method can be used to effectively introduce genes into target cells for the gene therapy of cancer and viral diseases, e.g. AIDS

XX

```
SQ Sequence 457 AA;
Query Match 100.0%; Score 2398; DB 2; Length 457;
Best Local Similarity 100.0%; Pred. No. 3.1e-172;
Matches 457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTDLRFTNIGDPTMRVTWAPPSIDLTFNFLVRYSPVKNEEDVAELSIKPSDNAVLTNLL 60
DB 1 PTDLRFTNIGDPTMRVTWAPPSIDLTFNFLVRYSPVKNEEDVAELSIKPSDNAVLTNLL 60
QY 61 PGTEYVSVSVVEQHESTPLRGOKTGLDPTGIDFSDITANSFVHIAPRATITGYR 120
DB 61 PGTEYVSVSVVEQHESTPLRGOKTGLDPTGIDFSDITANSFVHIAPRATITGYR 120
QY 121 IRHHEHFGSRPREDRVPHSRNSITLNTLTPGTEYVSVIVALNGREESPLLIGQOSTVSD 180
DB 121 IRHHEHFGSRPREDRVPHSRNSITLNTLTPGTEYVSVIVALNGREESPLLIGQOSTVSD 180
QY 181 VPRDLEVVAAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
DB 181 VPRDLEVVAAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
QY 241 PGVDYITTVYAVTGRGDSPASSKPIISINYTEIDKPSMAAGSITTLPALPEDGGSGAPPP 300
DB 241 PGVDYITTVYAVTGRGDSPASSKPIISINYTEIDKPSMAAGSITTLPALPEDGGSGAPPP 300
QY 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLOLQAEERGVSIGVCANRYL 360
DB 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLOLQAEERGVSIGVCANRYL 360
QY 361 AMKEDGRLLASKCVTDECFFERLESNNYRYSRKYTSWYVALKRTGYKLGSKTGPQQ 420
DB 361 AMKEDGRLLASKCVTDECFFERLESNNYRYSRKYTSWYVALKRTGYKLGSKTGPQQ 420
QY 421 KAILFLPMSAASDELPLVTLPHNPHLHGPEILDVPST 457
DB 421 KAILFLPMSAASDELPLVTLPHNPHLHGPEILDVPST 457

RESULT 2
AAR40160
ID AAR40160 standard; peptide; 432 AA.
XX
AC AAR40160;
XX
DT 07-FEB-1994 (first entry)
XX
DE Human FN/bFGF fusion peptide #1.
XX
KW Human; fibronectin; FN; fibroblast cell growth factor; FGF; fusion;
XX cell adhesion; cell growth; anti-aging; cosmetics; wound healing;
XX surgery.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..277
FT Peptide /note= "Human FN fragment 1239-1515"
FT Peptide 278..432
FT Peptide /note= "bFGF fragment"
XX
PN JP05178897-A.
XX
PD 20-JUL-1993.
XX
PF 05-MAR-1992; 92JP-00083220.
XX
PR 14-OCT-1991; 91JP-00291959.
XX
PA (TAKI ) TAKARA SHUZO CO LTD.
XX
DR WPI; 1993-261656/33.
XX

PT Synthetic functional polypeptide to promote wound healing, etc. - contg.
PT cell adhesion polypeptide from fibronectin and fibroblast growth factor
PT polypeptide, opt. linked by spacer.
XX
XX Disclosure; Page 7-8; 13pp; Japanese.
XX
CC The sequences given in AAR40158-63 represent human fibronectin (FN) and
CC fibroblast cell growth factor (FGF) fragments which were used in the
CC production of fusion polypeptides which are able to stimulate cell
CC adhesion and cell growth. these fusion peptides may be used for anti-
CC aging cosmetics and in wound healing after surgery
XX
SQ Sequence 432 AA;
Query Match 94.1%; Score 2257; DB 2; Length 432;
Best Local Similarity 99.8%; Pred. No. 1.2e-161;
Matches 431; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTDLRFTNIGDPTMRVTWAPPSIDLTFNFLVRYSPVKNEEDVAELSIKPSDNAVLTNLL 60
DB 1 PTDLRFTNIGDPTMRVTWAPPSIDLTFNFLVRYSPVKNEEDVAELSIKPSDNAVLTNLL 60
QY 61 PGTEYVSVSVVEQHESTPLRGOKTGLDPTGIDFSDITANSFVHIAPRATITGYR 120
DB 61 PGTEYVSVSVVEQHESTPLRGOKTGLDPTGIDFSDITANSFVHIAPRATITGYR 120
QY 121 IRHHEHFGSRPREDRVPHSRNSITLNTLTPGTEYVSVIVALNGREESPLLIGQOSTVSD 180
DB 121 IRHHEHFGSRPREDRVPHSRNSITLNTLTPGTEYVSVIVALNGREESPLLIGQOSTVSD 180
QY 181 VPRDLEVVAAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
DB 181 VPRDLEVVAAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
QY 241 PGVDYITTVYAVTGRGDSPASSKPIISINYTEIDKPSMAAGSITTLPALPEDGGSGAPPP 300
DB 241 PGVDYITTVYAVTGRGDSPASSKPIISINYTEIDKPSMAAGSITTLPALPEDGGSGAPPP 300
QY 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLOLQAEERGVSIGVCANRYL 360
DB 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLOLQAEERGVSIGVCANRYL 360
QY 361 AMKEDGRLLASKCVTDECFFERLESNNYRYSRKYTSWYVALKRTGYKLGSKTGPQQ 420
DB 361 AMKEDGRLLASKCVTDECFFERLESNNYRYSRKYTSWYVALKRTGYKLGSKTGPQQ 420
QY 421 KAILFLPMSAAS 432
DB 421 KAILFLPMSAKS 432

RESULT 3
AAR33339
ID AAR33339 standard; protein; 432 AA.
XX
AC AAR33339;
XX
DT 23-FEB-1998 (first entry)
XX
DE Human fibroblast growth factor oligopeptide.
XX
KW Human; fibroblast growth factor; FGF; target cell; transfection;
XX retroviral vector; gene therapy; cancer; viral disease;
XX acquired immunodeficiency syndrome; AIDS.
XX
OS Homo sapiens.
XX
PN WO9718318-A1.
XX
PD 22-MAY-1997.
XX
PF 07-NOV-1996; 96WO-JP003254.
XX
```

PR 13-NOV-1995; 95JP-00294382.  
PR 08-MAR-1996; 96JP-00051847.  
XX  
PA (TAKI ) TAKARA SHUZO CO LTD.  
XX  
PI Asada K, Uemori T, Ueno T, Koyama N, Hashino K, Kato I;  
XX  
XX WPI; 1997-289294/26.  
DR  
XX  
XX Method for increasing efficacy of gene transfer to target cell using  
PT retrovirus - by infection of the target cell in the presence of a  
PT substance which binds to the virus and a substance which binds to the  
PT target cell.  
XX  
PS Claim 42; Page 94-97; 194pp; Japanese.  
XX  
XX The present sequence is a human fibroblast growth factor (FGF)  
CC oligopeptide, which was used in the development of a novel method for  
CC increasing the efficiency of gene introduction into a target cell using a  
CC retroviral vector. The method comprises carrying out viral infection of  
CC the target cell in the presence of a retrovirus and target cell binding  
CC substance or substances. The method can be used to effectively introduce  
CC genes into target cells for the gene therapy of cancer and viral  
CC diseases, e.g. AIDS  
XX  
SQ Sequence 432 AA;  
Query Match 94.1%; Score 2257; DB 2; Length 432;  
Best Local Similarity 99.8%; Pred. No. 1.2e-161;  
Matches 431; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 PTDLRFNIGPDMRTVMTWAPPSSIDLTFNLVRYSPVKNEEDVAELISPSDNAVLTNLL 60  
DB 1 PTDLRFNIGPDMRTVMTWAPPSSIDLTFNLVRYSPVKNEEDVAELISPSDNAVLTNLL 60  
QY 61 PGTEYVSVSSVYEQHESTPLRGQKTGLDSTPGIDFSDITANSFTVHVIAPRATITGYR 120  
DB 61 PGTEYVSVSSVYEQHESTPLRGQKTGLDSTPGIDFSDITANSFTVHVIAPRATITGYR 120  
QY 121 IRHHPHFSGRPREDRVPHSRNSITLNTLPGTEYVSVISVALNGREESPLLIGQQSTVSD 180  
DB 121 IRHHPHFSGRPREDRVPHSRNSITLNTLPGTEYVSVISVALNGREESPLLIGQQSTVSD 180  
QY 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240  
DB 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240  
QY 241 PGVDYITVYAVTGRGDSPASSKPI SINRYRTEIDKPSMAAGSITTLPALPEDGGGAFPP 300  
DB 241 PGVDYITVYAVTGRGDSPASSKPI SINRYRTEIDKPSMAAGSITTLPALPEDGGGAFPP 300  
QY 301 GHFKDPKRLYCKNGGFFRLIHPDGRVDGVREKSDPHIKLQQAEEGVVSIKGVCANRYL 360  
DB 301 GHFKDPKRLYCKNGGFFRLIHPDGRVDGVREKSDPHIKLQQAEEGVVSIKGVCANRYL 360  
QY 361 AMKEDGRLLASKCVTDECFERLESNNYNTYRSRYTSYVALKRTGYKLGSKTGPQG 420  
DB 361 AMKEDGRLLASKCVTDECFERLESNNYNTYRSRYTSYVALKRTGYKLGSKTGPQG 420  
QY 421 KAILFLPMSAAS 432  
DB 421 KAILFLPMSAKS 432  
RESULT 4  
AA05457  
ID RAY05457 standard; protein; 432 AA.  
XX  
XX AA05457;  
XX  
XX 07-JUL-1999 (first entry)  
XX  
XX Fibronectin receptor targeting HIV strain C-FGF.A.

XX Fibronectin receptor; HIV; infection; therapy.  
XX  
XX Unidentified.  
XX JP10029952-A.  
XX  
XX 03-FEB-1998.  
XX  
XX 16-JUL-1996; 96JP-00185893.  
XX  
XX 16-JUL-1996; 96JP-00185893.  
XX  
XX (TAKI ) TAKARA SHUZO CO LTD.  
XX  
XX WPI; 1998-163674/15.  
XX  
XX Control of human immunodeficiency virus infection - using composition  
PT comprising replication defective HIV vector.  
XX  
XX Example 3; Page 21-22; 24pp; Japanese.  
XX  
XX This sequence represents a fibronectin receptor that can be used in the  
CC method of the invention. The method is for the control of human  
CC immunodeficiency virus (HIV) infection using a composition which  
CC comprises a functional substance which participates in the infection of  
CC HIV. The method is used to control HIV-infection  
XX  
SQ Sequence 432 AA;  
Query Match 94.1%; Score 2257; DB 2; Length 432;  
Best Local Similarity 99.8%; Pred. No. 1.2e-161;  
Matches 431; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 PTDLRFNIGPDMRTVMTWAPPSSIDLTFNLVRYSPVKNEEDVAELISPSDNAVLTNLL 60  
DB 1 PTDLRFNIGPDMRTVMTWAPPSSIDLTFNLVRYSPVKNEEDVAELISPSDNAVLTNLL 60  
QY 61 PGTEYVSVSSVYEQHESTPLRGQKTGLDSTPGIDFSDITANSFTVHVIAPRATITGYR 120  
DB 61 PGTEYVSVSSVYEQHESTPLRGQKTGLDSTPGIDFSDITANSFTVHVIAPRATITGYR 120  
QY 121 IRHHPHFSGRPREDRVPHSRNSITLNTLPGTEYVSVISVALNGREESPLLIGQQSTVSD 180  
DB 121 IRHHPHFSGRPREDRVPHSRNSITLNTLPGTEYVSVISVALNGREESPLLIGQQSTVSD 180  
QY 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240  
DB 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240  
QY 241 PGVDYITVYAVTGRGDSPASSKPI SINRYRTEIDKPSMAAGSITTLPALPEDGGGAFPP 300  
DB 241 PGVDYITVYAVTGRGDSPASSKPI SINRYRTEIDKPSMAAGSITTLPALPEDGGGAFPP 300  
QY 301 GHFKDPKRLYCKNGGFFRLIHPDGRVDGVREKSDPHIKLQQAEEGVVSIKGVCANRYL 360  
DB 301 GHFKDPKRLYCKNGGFFRLIHPDGRVDGVREKSDPHIKLQQAEEGVVSIKGVCANRYL 360  
QY 361 AMKEDGRLLASKCVTDECFERLESNNYNTYRSRYTSYVALKRTGYKLGSKTGPQG 420  
DB 361 AMKEDGRLLASKCVTDECFERLESNNYNTYRSRYTSYVALKRTGYKLGSKTGPQG 420  
QY 421 KAILFLPMSAAS 432  
DB 421 KAILFLPMSAKS 432  
RESULT 5  
ABP72002  
ID ABP72002 standard; protein; 432 AA.  
XX  
XX ABP72002;  
XX

DT 16-MAY-2003 (first entry)  
XX Human fibroblast growth factor C-FGF.  
DE Human; drug; cell adhesion; fibroblast growth factor; FGF; cardiant;  
KW osteopathic; angiogenesis promotion; arterial disease; cartilage damage;  
KW C-FGF.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX WO2003004066-A1.  
XX 16-JAN-2003.  
XX 04-JUL-2002; 2002WO-JP006772.  
XX 05-JUL-2001; 2001JP-00204300.  
XX (TAKA-) TAKARA BIO INC.  
XX Tanaka K, Chono H, Morishita M, Konishi H, Ueno M, Asada K;  
PI Kato I;  
PI  
DR WPI; 2003-210320/20.  
DR N-PSDB; ABZ82340.  
XX  
XX Gene therapeutics to provide fusion polypeptide with cell adhesion and  
PT fibroblast growth factor activities, applicable in angiogenesis promotion  
PT and treatment of arterial diseases and cartilage damage.  
XX  
PS Claim 8; Page 36-38; 53pp; Japanese.  
XX  
XX The invention relates to novel drugs containing a gene encoding a  
CC polypeptide which has the amino acid sequence of a polypeptide having a  
CC cell adhesion activity and another polypeptide with an amino acid  
CC sequence of a polypeptide having a fibroblast growth factor activity. The  
CC drugs of the invention have cardiant, and osteopathic activity. The gene  
CC therapeutics of the invention are applicable in angiogenesis promotion  
CC and treatment of arterial diseases and cartilage damage. The sequence  
CC represents the human C-FGF protein, a fusion of C277 and bFGF  
XX  
SQ Sequence 432 AA;  
  
Query Match 94.1%; Score 2257; DB 6; Length 432;  
Best Local Similarity 99.8%; Pred. No. 1.2e-161;  
Matches 431; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 PTDLRFNIGPDMRVTWAPPSSIDLTFNLFVRYSPVKNEEDVAELSIQSPSDNAVVLNLL 60  
DB 1 PTDLRFNIGPDMRVTWAPPSSIDLTFNLFVRYSPVKNEEDVAELSIQSPSDNAVVLNLL 60  
  
QY 61 PGTEYVSVSVYQHESTPLRGQKTGLDSTGIDFSDITANSFTVHWIAPRATITGYR 120  
DB 61 PGTEYVSVSVYQHESTPLRGQKTGLDSTGIDFSDITANSFTVHWIAPRATITGYR 120  
  
QY 121 IRHPEHFGSRPREDRVPHSRNSITLNLTPGTEYVSVIVALNGREESPLLIGQSTVSD 180  
DB 121 IRHPEHFGSRPREDRVPHSRNSITLNLTPGTEYVSVIVALNGREESPLLIGQSTVSD 180  
  
QY 181 VRDLEVAATTSLLISWDAPAVTVYRITTYGETGNSPVQETVPFGSKSTATISGLK 240  
DB 181 VRDLEVAATTSLLISWDAPAVTVYRITTYGETGNSPVQETVPFGSKSTATISGLK 240  
  
QY 241 PGVDYITVYATGTGDSPASSKPIISINVRTEIDKPSMAAGSITTLPALPEDGGSGAPP 300  
DB 241 PGVDYITVYATGTGDSPASSKPIISINVRTEIDKPSMAAGSITTLPALPEDGGSGAPP 300  
  
QY 301 GHFKDPKRLYCKNGGFFLRHDPGRVDGVREKSDPHIKLQQAERGVVSVKGVCANRYL 360  
DB 301 GHFKDPKRLYCKNGGFFLRHDPGRVDGVREKSDPHIKLQQAERGVVSVKGVCANRYL 360  
  
QY 361 AMKEDGRLLASKCVTDECFERLESNNYTRSRKYTSWYVALKRTGYKLGSKTGPQ 420

DB 361 AMKEDGRLLASKCVTDECFERLESNNYTRSRKYTSWYVALKRTGYKLGSKTGPQ 420  
QY 421 KAILFLPMSAAS 432  
DB 421 KAILFLPMSAKS 432  
  
RESULT 6  
ABP72004  
ID ABP72004 standard; protein; 455 AA.  
XX  
AC ABP72004;  
XX  
XX 16-MAY-2003 (first entry)  
XX Human fibroblast growth factor SC-FGF.  
DE Human; drug; cell adhesion; fibroblast growth factor; FGF; cardiant;  
KW osteopathic; angiogenesis promotion; arterial disease; cartilage damage;  
KW SC-FGF.  
XX Homo sapiens.  
XX WO2003004066-A1.  
XX 16-JAN-2003.  
XX 04-JUL-2002; 2002WO-JP006772.  
XX 05-JUL-2001; 2001JP-00204300.  
XX (TAKA-) TAKARA BIO INC.  
XX Tanaka K, Chono H, Morishita M, Konishi H, Ueno M, Asada K;  
PI Kato I;  
PI  
DR WPI; 2003-210320/20.  
DR N-PSDB; ABZ82342.  
XX  
XX Gene therapeutics to provide fusion polypeptide with cell adhesion and  
PT fibroblast growth factor activities, applicable in angiogenesis promotion  
PT and treatment of arterial diseases and cartilage damage.  
XX  
XX Claim 13; Page 41-43; 53pp; Japanese.  
XX  
XX The invention relates to novel drugs containing a gene encoding a  
CC polypeptide which has the amino acid sequence of a polypeptide having a  
CC cell adhesion activity and another polypeptide with an amino acid  
CC sequence of a polypeptide having a fibroblast growth factor activity. The  
CC drugs of the invention have cardiant, and osteopathic activity. The gene  
CC therapeutics of the invention are applicable in angiogenesis promotion  
CC and treatment of arterial diseases and cartilage damage. The sequence  
CC represents the human SC-FGF protein, a fusion of FGF-4 signal and C-FGF  
XX  
SQ Sequence 455 AA;  
  
Query Match 94.1%; Score 2257; DB 6; Length 455;  
Best Local Similarity 99.8%; Pred. No. 1.3e-161;  
Matches 431; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 PTDLRFNIGPDMRVTWAPPSSIDLTFNLFVRYSPVKNEEDVAELSIQSPSDNAVVLNLL 60  
DB 24 PTDLRFNIGPDMRVTWAPPSSIDLTFNLFVRYSPVKNEEDVAELSIQSPSDNAVVLNLL 83  
  
QY 61 PGTEYVSVSVYQHESTPLRGQKTGLDSTGIDFSDITANSFTVHWIAPRATITGYR 120  
DB 84 PGTEYVSVSVYQHESTPLRGQKTGLDSTGIDFSDITANSFTVHWIAPRATITGYR 143  
  
QY 121 IRHPEHFGSRPREDRVPHSRNSITLNLTPGTEYVSVIVALNGREESPLLIGQSTVSD 180  
DB 144 IRHPEHFGSRPREDRVPHSRNSITLNLTPGTEYVSVIVALNGREESPLLIGQSTVSD 203

QY 181 VPRDLVVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240  
Db 204 VPRDLVVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 263  
QY 241 PGVDYTIIVAVTGRGDSPASSKPIISINRYTEIDKPSMAAGSIITLTPALPEDGSGGAFPP 300  
Db 264 PGVDYTIIVAVTGRGDSPASSKPIISINRYTEIDKPSMAAGSIITLTPALPEDGSGGAFPP 323  
QY 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLQQAEEGWWSIKGVCANRYL 360  
Db 324 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLQQAEEGWWSIKGVCANRYL 383  
QY 361 AMKEDGRLASKCVTDECFERLESNNYNTYRSKYTSWYVALKRTGYKLGSKTGPQ 420  
Db 384 AMKEDGRLASKCVTDECFERLESNNYNTYRSKYTSWYVALKRTGYKLGSKTGPQ 443  
QY 421 KAILFLPMSAAS 432  
Db 444 KAILFLPMSAKS 455  
RESULT 7  
AAR40162  
ID AAR40162 standard; peptide; 432 AA.  
AC AAR40162;  
DT 07-FEB-1994 (first entry)  
XX Human FN/bFGF fusion peptide #2.  
XX Human; fibronectin; FN; fibroblast cell growth factor; FGF; fusion;  
KW cell adhesion; cell growth; anti-aging; cosmetics; wound healing;  
KW surgery.  
XX Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..277  
FT /note= "Human FN fragment 1239-1515"  
FT Peptide 278..432  
FT /note= "bFGF fragment"  
XX  
PN JP05178897-A.  
XX  
XX 20-JUL-1993.  
XX  
XX 05-MAR-1992; 92JP-00083220.  
XX  
XX 14-OCT-1991; 91JP-00291959.  
XX  
XX (TAKI ) TAKARA SHUZO CO LTD.  
XX  
XX WPI; 1993-261656/33.  
XX  
FT Synthetic functional polypeptide to promote wound healing, etc. - contg.  
FT cell adhesion polypeptide from fibronectin and fibroblast growth factor  
FT polypeptide, opt. linked by spacer.  
XX  
XX Disclosure; Page 9-10; 13pp; Japanese.  
XX  
CC The sequences given in AAR40158-63 represent human fibronectin (FN) and  
CC fibroblast cell growth factor (FGF) fragments which were used in the  
CC production of fusion polypeptides which are able to stimulate cell  
CC adhesion and cell growth. These fusion peptides may be used for anti-  
CC aging cosmetics and in wound healing after surgery  
XX  
SQ Sequence 432 AA;

Query Match 93.6%; Score 2245; DB 2; Length 432;  
Best Local Similarity 99.3%; Pred. No. 9.8e-161;  
Matches 429; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PTDLRNTIGPDMRTWAPPISIDLTNLFVRYSPVKNEEDVAELISPSDNAVLTNLL 60  
Db 1 PTDLRNTIGPDMRTWAPPISIDLTNLFVRYSPVKNEEDVAELISPSDNAVLTNLL 60  
QY 61 PGTEYVVSVSSVVEQHESTPLRGROKTGLDSTPGIDFSDITANSFTVHMTAPRATITGYR 120  
Db 61 PGTEYVVSVSSVVEQHESTPLRGROKTGLDSTPGIDFSDITANSFTVHMTAPRATITGYR 120  
QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVVSIVALNGREESPLLIQQOSTVSD 180  
Db 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVVSIVALNGREESPLLIQQOSTVSD 180  
QY 181 VPRDLVVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240  
Db 181 VPRDLVVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240  
QY 241 PGVDYTIIVAVTGRGDSPASSKPIISINRYTEIDKPSMAAGSIITLTPALPEDGSGGAFPP 300  
Db 241 PGVDYTIIVAVTGRGDSPASSKPIISINRYTEIDKPSMAAGSIITLTPALPEDGSGGAFPP 300  
QY 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLQQAEEGWWSIKGVCANRYL 360  
Db 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLQQAEEGWWSIKGVCANRYL 360  
QY 361 AMKEDGRLASKCVTDECFERLESNNYNTYRSKYTSWYVALKRTGYKLGSKTGPQ 420  
Db 361 AMKEDGRLASKCVTDECFERLESNNYNTYRSKYTSWYVALKRTGYKLGSKTGPQ 420  
QY 421 KAILFLPMSAAS 432  
Db 421 KAILFLPMSAKS 432  
RESULT 8  
AAR08044  
ID AAR08044 standard; protein; 574 AA.  
XX  
AC AAR08044;  
XX  
DT 25-MAR-2003 (revised)  
DT 25-FEB-1991 (first entry)  
XX  
XX Fibronectin deriv. for inhibition of angiogenesis.  
XX  
XX Fibronectin; heparin; angiogenesis; metastasis; tumours; prolasias;  
KW contraceptive; retinopathy; osteoporosis; rheumatism.  
XX  
XX Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..277  
FT /label= FN fragment of cell-binding domain  
FT Peptide 279..549  
FT /label= FN fragment of heparin-binding domain  
XX  
XX EP399806-A.  
PN  
XX  
PD 28-NOV-1990.  
XX  
XX 23-MAY-1990; 90EP-00305619.  
XX  
XX 26-MAY-1989; 89JP-00131453.  
PR 01-DEC-1989; 89JP-00310536.  
XX  
XX (TAKI ) TAKARA SHUZO CO LTD.  
PA  
XX Taguchi Y, Ohdate Y, Kawase Y, Goto S, Fusao K, Kato I, Saiki I;  
PI Azuma I;  
XX  
XX WPI; 1990-356419/48.  
XX  
XX Fibronectin deriv. - contg. cell binding and heparin-binding domains  
FT linked via aminoacid residue or peptide, used as inhibitor of



```
QY 399 -----SWYVALKRTGQY-----KLGSK-----TG- 417
Db 462 DAPSNLRLATTNSLLVSWQPPRARITGYIIKYEKGSPPREVVRPRPGVTEATITGL 521
QY 418 -PGOKAILFL-----PM--SAASDELPLVTLPHNPHNLHGPEILDVPST 457
Db 522 EPGTEYTIYVIALKNNQKSEPLIGRKKTDLPQLVTLPHNPHNLHGPEILDVPST 574

RESULT 10
AAW33349
ID AAW33349 standard; protein; 574 AA.
XX
AC AAW33349;
XX
DT 23-FEB-1998 (first entry)
XX
DE Oligopeptide CH-296.
XX
KW Oligopeptide CH-296; target cell; transfection; retroviral vector;
KW gene therapy; cancer; viral disease; acquired immunodeficiency syndrome;
KW AIDS.
XX
OS Synthetic.
XX
PN W09718318-A1.
XX
PD 22-MAY-1997.
XX
PF 07-NOV-1996; 96WO-JP003254.
XX
PR 13-NOV-1995; 95JP-00294382.
PR 08-MAR-1996; 96JP-00051847.
XX
PA (TAKI ) TAKARA SHUZO CO LTD.
XX
PI Asada K, Uemori T, Ueno T, Koyama N, Hashino K, Kato I;
XX
DR WPI; 1997-289294/26.
XX
PT Method for increasing efficacy of gene transfer to target cell using
PT retrovirus - by infection of the target cell in the presence of a
PT substance which binds to the virus and a substance which binds to the
PT target cell.
XX
PS Disclosure; Page 130-133; 194pp; Japanese.
XX
CC The present sequence is the oligopeptide CH-296, which was used in the
CC development of a novel method for increasing the efficiency of gene
CC introduction into a target cell using a retroviral vector. The method
CC comprises carrying out viral infection of the target cell in the presence
CC of a retrovirus and target cell binding substance or substances. The
CC method can be used to effectively introduce genes into target cells for
CC the gene therapy of cancer and viral diseases, e.g. AIDS
XX
SQ Sequence 574 AA;

Query Match 64.1%; Score 1536.5; DB 2; Length 574;
Best Local Similarity 58.0%; Pred. No. 3.5e-107;
Matches 34%; Conservative 26; Mismatches 68; Indels 155; Gaps 14;

QY 1 PTDLRFNIGDMRVMTWAPPSPIDLTNFLVRYSPVKNEEDVAELSPSDNAVLNLL 60
Db 1 PTDLRFNIGDMRVMTWAPPSPIDLTNFLVRYSPVKNEEDVAELSPSDNAVLNLL 60
QY 61 PGTEYVSVSSVVEQHESTPLRGQKTGLDPSPTGIDFSDITANSFTVHWIAPRATITGYR 120
Db 61 PGTEYVSVSSVVEQHESTPLRGQKTGLDPSPTGIDFSDITANSFTVHWIAPRATITGYR 120
QY 121 IRHHPHFSGRPREDRVPHSRNSTLTNLTGTEYVSVIVALNGREESPLLIGQQSTVSD 180
Db 121 IRHHPHFSGRPREDRVPHSRNSTLTNLTGTEYVSVIVALNGREESPLLIGQQSTVSD 180
```

```
QY 181 VPRDLEVAATPTSLILISWDAPAVTVYRITYTGETGNSPVQEFVTPGSKSTATISGLK 240
Db 181 VPRDLEVAATPTSLILISWDAPAVTVYRITYTGETGNSPVQEFVTPGSKSTATISGLK 240
QY 241 PGVDYTIIVAVTGRGDSPASSKPIISINVRTEIDKPSMAAGSITTL--PALPEDGGSGA 297
Db 241 PGVDYTIIVAVTGRGDSPASSKPIISINVRTEIDKPSMAIPAPTDLKFQVTTPTSLSAQW 300
QY 298 FPEGHFKDPKRLYCKNGGPFRLIHPDGRVDGVREKSDPHIKLQLOAEEGVWSIKGYCAN 357
Db 301 TPN-----VLTGYRVVTP-----KKTGPMKEINLAPDSSSVVWVGLMVA 344
QY 358 RY-----LANKEDGRLLASK-----CVTDEC----- 378
Db 345 KYEVSIVALKD---TLTSRPAQGVVTTLENVSPRRARVTDATETTTITISWRTKTETITG 401
QY 379 FFERLESNNYNTYRS-----RKYT----- 398
Db 402 FQVDVAPANGQTPIQRTIKPDVRSYITIGLQPGDYKIYLYTLNDNARSPVVIDASTAI 461
QY 399 -----SWYVALKRTGQY-----KLGSK-----TG- 417
Db 462 DAPSNLRLATTNSLLVSWQPPRARITGYIIKYEKGSPPREVVRPRPGVTEATITGL 521
QY 418 -PGOKAILFL-----PM--SAASDELPLVTLPHNPHNLHGPEILDVPST 457
Db 522 EPGTEYTIYVIALYNNQKSEPLIGRKKTDLPQLVTLPHNPHNLHGPEILDVPST 574

RESULT 11
AAW13572
ID AAW13572 standard; protein; 574 AA.
XX
AC AAW13572;
XX
DT 14-NOV-1997 (first entry)
XX
DE Escherichia coli fibronectin protein.
KW Immunology; haematopoietic somatic cell; vertebrate; mammal; bird;
KW amphibian; fish; reptile; mouse; human; bone marrow.
XX
OS Escherichia coli.
XX
PN W09707670-A1.
XX
PD 06-MAR-1997.
XX
PF 19-AUG-1996; 96WO-JP002312.
XX
PR 30-AUG-1995; 95JP-00221583.
XX
PA (TAKI ) TAKARA SHUZO CO LTD.
XX
PI Matsushita H, Kato I;
XX
DR WPI; 1997-178821/16.
XX
PT Vertebrate immunologically tolerant to foreign proteins - produced by
PT transforming haematopoietic somatic cells with a gene encoding foreign
PT protein.
XX
PS Disclosure; Page 34-37; 48pp; Japanese.
XX
CC A method has been developed for producing a vertebrate immunologically
CC tolerant to foreign proteins. The method involves: (a) transforming
CC haematopoietic somatic cells of a vertebrate with a gene encoding foreign
CC proteins, (b) producing an immunodeficient vertebrate, and (c)
CC transferring the transformed cells into the immunodeficient vertebrate.
CC The present sequence represents an Escherichia coli fibronectin protein
CC which can be used as an example of the above method. The vertebrate
CC may be used in the production of foreign proteins with pharmaceutical
CC applications. They may also have potential uses as models for human
```

CC	diseases	DR	WPI; 1998-163674/15.
XX		XX	Control of human immunodeficiency virus infection - using composition comprising replication defective HIV vector.
XX		XX	Claim 3; Page 11-14; 24pp; Japanese.
XX		XX	This sequence represents a fibronectin receptor that can be used in the method of the invention. The method is for the control of human immunodeficiency virus (HIV) infection using a composition which comprises a functional substance which participates in the infection of HIV. The method is used to control HIV-infection
XX		XX	Sequence 574 AA;
XX		XX	Query Match 64.1%; Score 1536.5; DB 2; Length 574;
XX		XX	Best Local Similarity 58.0%; Pred. No. 3.5e-107;
XX		XX	Matches 344; Conservative 26; Mismatches 68; Indels 155; Gaps 14;
QY	1 PTDLRFNIGPDTMRVTWAPPSSIDLTNPLVRYSPVKNEEDVAELISPSDNAVLTNLL 60	QY	1 PTDLRFNIGPDTMRVTWAPPSSIDLTNPLVRYSPVKNEEDVAELISPSDNAVLTNLL 60
Db	1 PTDLRFNIGPDTMRVTWAPPSSIDLTNPLVRYSPVKNEEDVAELISPSDNAVLTNLL 60	Db	1 PTDLRFNIGPDTMRVTWAPPSSIDLTNPLVRYSPVKNEEDVAELISPSDNAVLTNLL 60
QY	61 PGTEYVSVSSVYEQHESTPLRGQKTGLDPTGIDFSDITANSFTVHMIAPRATITGYR 120	QY	61 PGTEYVSVSSVYEQHESTPLRGQKTGLDPTGIDFSDITANSFTVHMIAPRATITGYR 120
Db	61 PGTEYVSVSSVYEQHESTPLRGQKTGLDPTGIDFSDITANSFTVHMIAPRATITGYR 120	Db	61 PGTEYVSVSSVYEQHESTPLRGQKTGLDPTGIDFSDITANSFTVHMIAPRATITGYR 120
QY	121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGREESPLLIGQOSTVSD 180	QY	121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGREESPLLIGQOSTVSD 180
Db	121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGREESPLLIGQOSTVSD 180	Db	121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGREESPLLIGQOSTVSD 180
QY	181 VPRDLEVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240	QY	181 VPRDLEVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
Db	181 VPRDLEVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240	Db	181 VPRDLEVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
QY	241 PGVDYITIVYAVTGRGDSPASSKPIINRYRTEIDKPSMAAGSITTL---PALPEDGGSGA 297	QY	241 PGVDYITIVYAVTGRGDSPASSKPIINRYRTEIDKPSMAAGSITTL---PALPEDGGSGA 297
Db	241 PGVDYITIVYAVTGRGDSPASSKPIINRYRTEIDKPSMAAGSITTL---PALPEDGGSGA 297	Db	241 PGVDYITIVYAVTGRGDSPASSKPIINRYRTEIDKPSMAAGSITTL---PALPEDGGSGA 297
QY	298 FPPGHFKDPKRLCKNGGFFLRHPDGRVDGVRKSDPHIKLOLQAEERGVSIGVCAN 357	QY	298 FPPGHFKDPKRLCKNGGFFLRHPDGRVDGVRKSDPHIKLOLQAEERGVSIGVCAN 357
Db	301 TPNP-----VQUTGYRVVTP-----KEKTGPMKEINLAPDSSVVVSGLMVAT 344	Db	301 TPNP-----VQUTGYRVVTP-----KEKTGPMKEINLAPDSSVVVSGLMVAT 344
QY	358 RY----LAMKEDGRLLASK-----CVTDEC----- 378	QY	358 RY----LAMKEDGRLLASK-----CVTDEC----- 378
Db	345 KYEVSVALKD---TLTSRPAQGVVTTLENVSPPRARVTDATETITISWRTKTETITG 401	Db	345 KYEVSVALKD---TLTSRPAQGVVTTLENVSPPRARVTDATETITISWRTKTETITG 401
QY	379 FFERLESNNYTRS-----RKYT----- 398	QY	379 FFERLESNNYTRS-----RKYT----- 398
Db	402 FQVDAVPANGQTPIQRTIKPDVRSYITITGLQPGTDYKIYLYTLNDNARSSPVVIDASTAI 461	Db	402 FQVDAVPANGQTPIQRTIKPDVRSYITITGLQPGTDYKIYLYTLNDNARSSPVVIDASTAI 461
QY	399 -----SWYVALKRTGOY-----KLGSK-----TG- 417	QY	399 -----SWYVALKRTGOY-----KLGSK-----TG- 417
Db	462 DAPSNLRFATPNSLLVSWQPPRARITGYIIKYEKPGSPREVVRPRPGVTEATITGL 521	Db	462 DAPSNLRFATPNSLLVSWQPPRARITGYIIKYEKPGSPREVVRPRPGVTEATITGL 521
QY	418 -PGQKAILFL-----PM--SAASDELQVLTLPHPNLHGPPEILDVPST 457	QY	418 -PGQKAILFL-----PM--SAASDELQVLTLPHPNLHGPPEILDVPST 457
Db	522 EPGTEYTIIVIALKNNKQSEPLIGRKKTKDELQVLTLPHPNLHGPPEILDVPST 574	Db	522 EPGTEYTIIVIALKNNKQSEPLIGRKKTKDELQVLTLPHPNLHGPPEILDVPST 574
XX		XX	RESULT 13
XX		XX	AAW97357
XX		XX	ID AAW97357 standard; peptide; 574 AA.
XX		XX	AAW97357;
XX		XX	12-MAY-1999 (first entry)
XX		XX	Protein sequence of the specification.
XX		XX	Retrovirus; gene transfer; serum-free medium; AIDS; cancer; leukaemia;
XX		XX	gene therapy.



OS Homo sapiens.  
FN WO9905301-A1.  
XX 04-FEB-1999.  
PD 15-JUL-1998; 98WO-JP003173.  
PF 23-JUL-1997; 97JP-00196772.  
XX (TAKI ) TAKARA SHUZO CO LTD.  
PA Bagnis C, Imbert A, Mannoni P;  
PI WPI; 1999-142951/12.  
DR Gene transfer by retrovirus in medium containing functional substance and  
XX optionally low-density lipoprotein - useful in medical sciences, cell and  
XX gene engineering, particularly for treating AIDS and cancers.  
XX Claim 5; Page 27-30; 32pp; Japanese.  
PS The specification describes a method for transferring a gene into target  
XX cells by a retrovirus using a serum-free medium. The culture medium of  
XX the target cells is serum free and contains an effective amount of a  
XX functional substance to elevate the gene transfer efficiency when both  
XX the retrovirus and target cells are present together. The gene transfer  
XX method is useful in medical sciences, cell engineering and genetic  
XX engineering, such as in the treatment of AIDS and cancers e.g. leukaemia  
XX by gene therapy  
XX  
SQ Sequence 574 AA;  
  
Query Match 64.1%; Score 1536.5; DB 2; Length 574;  
Best Local Similarity 58.0%; Pred. No. 3.5e-107;  
Matches 344; Conservative 26; Mismatches 68; Indels 155; Gaps 14;  
  
Qy 1 PTDLRFNIGDPTMRVTWAPPSIDLTNLFVRYSPVKNEEDVAELISPSDNAVLTNLL 60  
Db |||||  
Qy 1 PTDLRFNIGDPTMRVTWAPPSIDLTNLFVRYSPVKNEEDVAELISPSDNAVLTNLL 60  
Db |||||  
Qy 61 PGTEYVSVSSVVEQHESTPLRGOKTGLDPTGIDFSDITANSFTVHWIAPRATITGYR 120  
Db |||||  
Qy 61 PGTEYVSVSSVVEQHESTPLRGOKTGLDPTGIDFSDITANSFTVHWIAPRATITGYR 120  
Db |||||  
Qy 121 IRHHPHFSGRPREDRVPHSRNSITLNLTPGTEYVSVIVALNGREESPLLIGQQSTVSD 180  
Db |||||  
Qy 121 IRHHPHFSGRPREDRVPHSRNSITLNLTPGTEYVSVIVALNGREESPLLIGQQSTVSD 180  
Db |||||  
Qy 181 VPRDLEVAATPTSLISWDAPAVTVRYRYITYGETGNSPVQEFVPGSKSTATISGLK 240  
Db |||||  
Qy 181 VPRDLEVAATPTSLISWDAPAVTVRYRYITYGETGNSPVQEFVPGSKSTATISGLK 240  
Db |||||  
Qy 241 PGVDYITTVYAVTGRGDSPASSKPIISINYTEIDKPSMAIGSITTL---PALPEDGSGA 297  
Db |||||  
Qy 241 PGVDYITTVYAVTGRGDSPASSKPIISINYTEIDKPSMAIPAPTDLKTQVTPTSLSAQW 300  
Db |||||  
Qy 298 FPGCHFKDPKRLCYCKNGGFFLRHDPGRVDGVRKESDPHIKLOQAEERGSWSIKGCAN 357  
Db |||||  
Qy 301 TPN-----VOLTGYRVTP-----KEKTPMKKEINLAPDSSVVVSGLMVAT 344  
Db |||||  
Qy 358 RY----LAMKEDGRLLASK-----CVTDEC----- 378  
Db |||||  
Qy 345 KYEVSVVALKD---TLTSRAQGVVTTLENVSPRRARVDTATETTITISWRKTTETITG 401  
Db |||||  
Qy 379 FFERLESNNYTRS-----RKYT----- 398  
Db |||||  
Qy 402 FOVDAPVANGQTPQRTIKPDVRSYITGLQPGTDYKIVLYTLNDNARSPVVIDASTAI 461  
Db |||||  
Qy 399 -----SNVALKRTGOY-----KLGSK-----TG- 417  
Db |||||  
Qy 462 DAPSNRFLATNSLLVSHQPPRARITGVIKYEKPGSPREVPVRPPGVTEATITGL 521  
Db |||||

Qy 418 -PCQKAILFL-----PM--SAASDELQVLTLPHPNLHGPEILDVPST 457  
Db |||||  
Qy 522 EPGTEYTIYVIALKNQKSEPLIGRKTKDELQVLTLPHPNLHGPEILDVPST 574  
Db |||||  
  
RESULT 14  
ABP59436  
ID ABP59436 standard; protein; 574 AA.  
XX AC ABP59436;  
XX 09-JUN-2003 (first entry)  
XX Human fibronectin fragment CH-296, SEQ ID 6.  
XX Human; fibronectin; antibacterial; virucide; cytostatic; antiallergic;  
XX cytotoxic T-cell; CTL; cytotoxicity; CD44; cancer; infection.  
XX Homo sapiens.  
XX WO2003016511-A1.  
XX 27-FEB-2003.  
XX 15-AUG-2002; 2002WO-JP008298.  
XX 15-AUG-2001; 2001JP-00246747.  
XX 11-DEC-2001; 2001JP-00376966.  
XX 25-MAR-2002; 2002JP-00084428.  
XX (TAKA-) TAKARA BIO INC.  
XX Sagawa H, Ideno M, Kato I;  
XX WPI; 2003-289979/28.  
XX Induction of antigen-specific cytotoxic T cells by culture in presence of  
XX a substance binding to CD44 for production of therapeutic cells treating  
XX infectious and allergic disorders and cancer.  
XX Disclosure; Page 102-106; 142pp; Japanese.  
XX  
XX The present invention relates to a method for inducing cytotoxic T-cells  
XX (CTL) having antigen-specific cytotoxicity. The method comprises  
XX incubating cells capable of differentiating into CTL with a cellular or  
XX non-cellular antigen in the presence of a substance having one of the  
XX following properties: (1) binding to CD44; (2) regulating the signal  
XX generated by CD44 binding to its ligand; (3) inhibiting binding of a  
XX growth factor to its receptor; (4) regulating the signal generated by  
XX growth factor binding to its receptor; (5) containing fibronectin and/or  
XX its fragments. Antigen-specific CTL are used for the treatment of cancer  
XX and bacterial and viral infections. The present sequence is a fragment of  
XX human fibronectin, which was used to illustrate the method of the  
XX invention  
XX  
SQ Sequence 574 AA;  
  
Query Match 64.1%; Score 1536.5; DB 6; Length 574;  
Best Local Similarity 58.0%; Pred. No. 3.5e-107;  
Matches 344; Conservative 26; Mismatches 68; Indels 155; Gaps 14;  
  
Qy 1 PTDLRFNIGDPTMRVTWAPPSIDLTNLFVRYSPVKNEEDVAELISPSDNAVLTNLL 60  
Db |||||  
Qy 61 PGTEYVSVSSVVEQHESTPLRGOKTGLDPTGIDFSDITANSFTVHWIAPRATITGYR 120  
Db |||||  
Qy 61 PGTEYVSVSSVVEQHESTPLRGOKTGLDPTGIDFSDITANSFTVHWIAPRATITGYR 120  
Db |||||  
Qy 121 IRHHPHFSGRPREDRVPHSRNSITLNLTPGTEYVSVIVALNGREESPLLIGQQSTVSD 180  
Db |||||  
Qy 121 IRHHPHFSGRPREDRVPHSRNSITLNLTPGTEYVSVIVALNGREESPLLIGQQSTVSD 180  
Db |||||

QY 181 VPRDLEVAATPTSLLSWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240  
Db 181 VPRDLEVAATPTSLLSWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240  
QY 241 PGVDYTIIVAVTGRGDSPPASSKPIISINYTEIDKPSMAAGSITTL---PALPEDGSGA 297  
Db 241 PGVDYTIIVAVTGRGDSPPASSKPIISINYTEIDKPSMAIPAPTDLKTQVTPTSLSAQW 300  
QY 298 FPGGHFKDPKRLYCKNGGFFLRHPDGRVDGVRKSDPHIKLQQAEEGWVSIKGVCAN 357  
Db 301 TPPN-----VOLTGYRVVTP-----KEKTPMKKEINLAPDSSVVVSGLMVAT 344  
QY 358 RY----LAWKEDGRLLASK-----CVTDEC----- 378  
Db 345 KYEVSVALKD---TLTSRPAQGVVTTLENVSPRRARVTDATETITISWRTKTETITG 401  
QY 379 FFERLESNNYTYRS-----RKYT----- 398  
Db 402 FOVDVAVPANGOTPIORTIKPDVRSYITITGLQPGTDYKIYLYTLNDNARSPVVIDASTAI 461  
QY 399 -----SWYVALKRTQY-----KLGSK-----TG- 417  
Db 462 DAPSNLRFATTPNSLLVSWQPPRARITGYIIKYEKPGSPPREVVRPRPGVTEATITGL 521  
QY 418 -PGQKAILFL-----PM--SAASDELQVLTLPHPNLHGPEILDVPST 457  
Db 522 EPGTEYTIIVIALKNNQKSEPLIGRKKTDLPQLVTLPHNHLHGPEILDVPST 574

RESULT 15

ID ADD49015 standard; protein; 574 AA.  
XX AC  
XX ADD49015;  
XX DT 15-JAN-2004 (first entry)  
XX DE Fibronectin fragment CH-296 #SEQ ID 12.  
XX KW Immunomodulator; cytotoxic lymphocyte; fibronectin; CD8 positive cell;  
XX KW adoptive immunotherapy.  
XX OS Synthetic.  
XX PN WO2003080817-A1.  
XX PD 02-OCT-2003.  
XX PF 25-MAR-2003; 2003WO-JP003575.  
XX PR 25-MAR-2002; 2002JP-00084414.  
XX PA (TAKI ) TAKARA BIO INC.  
XX PI Sagawa H, Ideno M, Kato I;  
XX WPI; 2003-865122/80.

XX Method for producing cytotoxic lymphocytes, useful in adoptive  
PT immunotherapy, comprises culturing in the presence of fibronectin or its  
PT fragments.

XX Claim 11; SEQ ID NO 12; 131pp; Japanese.

XX The invention relates to a method for producing cytotoxic lymphocytes  
CC comprising induction, maintenance or dilution (dilution) and culturing  
CC them in the presence of fibronectin and/or its fragments. Also disclosed  
CC is a method for increasing the ratio of CD8 positive cells. The  
CC fibronectin fragments are preferably optionally mutated sequences  
CC comprising fully defined sequences of between 25 and 99 amino acids.  
CC Cytotoxic lymphocytes of the invention are useful in adoptive  
CC immunotherapy. In an example from the invention a 15 day old culture of  
CC LAK cells was assayed to determine the ratio of CD8 positive cells. The

CC results showed that those cells subjected to fibronectin fragments and  
CC anti CD3 antibody during the early and middle phase of induction had a  
CC higher ratio of CD8 positive cells than in the control cells. The current  
CC sequence represents a fibronectin fragment of the invention.

XX Sequence 574 AA;

Query Match 64.1%; Score 1536.5; DB 7; Length 574;  
Best Local Similarity 58.0%; Pred. No. 3 5e-107;  
Matches 344; Conservative 26; Mismatches 68; Indels 155; Gaps 14;

QY 1 PTDLFTNIGPTDMRTWAPPESIDLTNFLVRYSPVQNEEDVAELSIQSDNAVLTNLL 60  
Db 1 PTDLFTNIGPTDMRTWAPPESIDLTNFLVRYSPVQNEEDVAELSIQSDNAVLTNLL 60  
QY 61 PCTEYVWSSVSYQHESTPLGRQKTGLDSTGIDFSDITANSFTVHWIAPRAITGYR 120  
Db 61 PCTEYVWSSVSYQHESTPLGRQKTGLDSTGIDFSDITANSFTVHWIAPRAITGYR 120  
QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTPTCEYVWSSIVALNGREESPLLIQOQSTVSD 180  
Db 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTPTCEYVWSSIVALNGREESPLLIQOQSTVSD 180  
QY 181 VPRDLEVAATPTSLLSWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240  
Db 181 VPRDLEVAATPTSLLSWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240  
QY 241 PGVDYTIIVAVTGRGDSPPASSKPIISINYTEIDKPSMAAGSITTL---PALPEDGSGA 297  
Db 241 PGVDYTIIVAVTGRGDSPPASSKPIISINYTEIDKPSMAIPAPTDLKTQVTPTSLSAQW 300  
QY 298 FPGGHFKDPKRLYCKNGGFFLRHPDGRVDGVRKSDPHIKLQQAEEGWVSIKGVCAN 357  
Db 301 TPPN-----VOLTGYRVVTP-----KEKTPMKKEINLAPDSSVVVSGLMVAT 344  
QY 358 RY----LAWKEDGRLLASK-----CVTDEC----- 378  
Db 345 KYEVSVALKD---TLTSRPAQGVVTTLENVSPRRARVTDATETITISWRTKTETITG 401  
QY 379 FFERLESNNYTYRS-----RKYT----- 398  
Db 402 FOVDVAVPANGOTPIORTIKPDVRSYITITGLQPGTDYKIYLYTLNDNARSPVVIDASTAI 461  
QY 399 -----SWYVALKRTQY-----KLGSK-----TG- 417  
Db 462 DAPSNLRFATTPNSLLVSWQPPRARITGYIIKYEKPGSPPREVVRPRPGVTEATITGL 521  
QY 418 -PGQKAILFL-----PM--SAASDELQVLTLPHPNLHGPEILDVPST 457  
Db 522 EPGTEYTIIVIALKNNQKSEPLIGRKKTDLPQLVTLPHNHLHGPEILDVPST 574

Search completed: May 3, 2004, 13:09:44  
Job time : 75.4946 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 13:07:19 ; Search time 23.4541 Seconds  
(without alignments)  
1005.924 Million cell updates/sec

Title: US-09-775-964-5  
Perfect score: 2398  
Sequence: 1 PDLRTNIGPDMRTWAP.....LVTLPHPNLHGPILDVPST 457

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA: \*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2398	100.0	457	4	US-09-366-009-5
2	2398	100.0	457	4	US-08-809-156B-5
3	2257	94.1	432	1	US-07-959-369-8
4	2257	94.1	432	2	US-08-836-854-20
5	2257	94.1	432	4	US-09-366-009-4
6	2257	94.1	432	4	US-08-809-156B-4
7	2254	94.0	432	1	US-07-959-369-9
8	1536.5	64.1	574	2	US-08-836-854-21
9	1536.5	64.1	574	3	US-09-463-296-1
10	1536.5	64.1	574	4	US-09-366-009-24
11	1536.5	64.1	574	4	US-08-809-156B-24
12	1527	63.7	489	4	US-09-366-009-8
13	1527	63.7	489	4	US-08-809-156B-8
14	1499	62.5	2231	1	US-08-153-799-16
15	1485.5	61.9	302	2	US-08-836-854-5
16	1485.5	61.9	302	4	US-09-366-009-29
17	1485.5	61.9	302	4	US-08-809-156B-29
18	1474	61.5	2324	1	US-08-283-857-1
19	1474	61.5	2324	5	PCT-US95-09819-1
20	1474	61.5	2327	6	5455158-1
21	1474	61.5	2386	2	US-09-016-366A-12
22	1474	61.5	2446	2	US-08-551-356-2
23	1474	61.5	2446	5	PCT-US93-12687-2
24	1463.5	61.0	472	4	US-09-366-009-21
25	1463.5	61.0	472	4	US-08-809-156B-21
26	1463.5	61.0	549	1	US-08-836-854-11
27	1463.5	61.0	549	4	US-09-366-009-23

28	1463.5	61.0	549	4	US-08-809-156B-23	Sequence 23, Appl
29	1455	60.7	422	2	US-08-836-854-12	Sequence 12, Appl
30	1454	60.6	826	4	US-09-366-009-14	Sequence 14, Appl
31	1454	60.6	826	4	US-08-809-156B-14	Sequence 14, Appl
32	1453	60.6	446	2	US-08-836-854-15	Sequence 15, Appl
33	1450	60.5	332	2	US-08-836-854-13	Sequence 13, Appl
34	1448.5	60.4	464	2	US-08-836-854-19	Sequence 19, Appl
35	1448.5	60.4	464	4	US-09-366-009-7	Sequence 7, Appl
36	1448.5	60.4	464	4	US-08-809-156B-7	Sequence 9, Appl
37	1447.5	60.4	474	2	US-08-836-854-9	Sequence 17, Appl
38	1447	60.3	368	2	US-08-836-854-17	Sequence 14, Appl
39	1442	60.1	341	2	US-08-836-854-14	Sequence 18, Appl
40	1442	60.1	367	2	US-08-836-854-18	Sequence 22, Appl
41	1442	60.1	457	2	US-08-836-854-16	Sequence 22, Appl
42	1442	60.1	457	4	US-09-366-009-22	Sequence 3, Appl
43	1442	60.1	457	4	US-08-809-156B-22	Sequence 11, Appl
44	1437	59.9	277	1	US-07-959-369-3	
45	1437	59.9	279	1	US-07-959-369-11	

ALIGNMENTS

RESULT 1  
US-09-366-009-5  
; Sequence 5, Application US/09366009  
; Patent No. 6426042  
; GENERAL INFORMATION:  
; APPLICANT: Asada, Kiyozo  
; Uemori, Takashi  
; Koyama, No. 6426042uto  
; Hashino, Kimikazu  
; Kato, Ikunoshin  
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET  
; CELLS WITH RETROVIRUS  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: WEISER & ASSOCIATES  
; STREET: 230 South Fifteenth Street, Suite 500  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19102  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/366,009  
; FILING DATE: 02-Aug-1999  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/809,156  
; FILING DATE: <Unknown>  
; APPLICATION NUMBER: JP 294382/1995  
; FILING DATE: 13-NOV-1995  
; APPLICATION NUMBER: JP 051847/1996  
; FILING DATE: 08-MAR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weiser, Gerard J.  
; REGISTRATION NUMBER: 19,763  
; REFERENCE/DOCKET NUMBER: 977.6507P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-875-8383  
; TELEFAX: 215-875-8394  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 457 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: <Unknown>  
; TOPOLOGY: linear

```
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-366-009-5

Query Match          100.0%; Score 2398; DB 4; Length 457;
Best Local Similarity 100.0%; Pred. No. 1.2e-171;
Matches 457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTDLRFNIGPDMRVMTWAPPSSIDLTLNLFVRYSPVKNEEDVAELISPSDNAVLTNLL 60
Db 1 PTDLRFNIGPDMRVMTWAPPSSIDLTLNLFVRYSPVKNEEDVAELISPSDNAVLTNLL 60

QY 61 PGTEYVSVSSVYEQHESTPLRGQKTGLDSTPGIDFSDITANSFTVHWIAPRATITGYR 120
Db 61 PGTEYVSVSSVYEQHESTPLRGQKTGLDSTPGIDFSDITANSFTVHWIAPRATITGYR 120

QY 121 IRHHPHFSGRPREDRVPHSRNSITLNLTPGTEYVSVIVALNGREESPLLIGQOSTVSD 180
Db 121 IRHHPHFSGRPREDRVPHSRNSITLNLTPGTEYVSVIVALNGREESPLLIGQOSTVSD 180

QY 181 VPRDLEVVAAATPTSLISWDAPAVTVRYRITYTGEGNSPVQEFVPGSKSTATISGLK 240
Db 181 VPRDLEVVAAATPTSLISWDAPAVTVRYRITYTGEGNSPVQEFVPGSKSTATISGLK 240

QY 241 PGVDYTIITVAVTGRGDSPASSKPIISINRYRTEIDKPSMAAGSITTLPALPEDGGSGAFP 300
Db 241 PGVDYTIITVAVTGRGDSPASSKPIISINRYRTEIDKPSMAAGSITTLPALPEDGGSGAFP 300

QY 301 GHFKDPKRLYCKNGGFFLRIHPDGRVDGVREKSDPHIKLOQAEERGVSIVKGVCANRYL 360
Db 301 GHFKDPKRLYCKNGGFFLRIHPDGRVDGVREKSDPHIKLOQAEERGVSIVKGVCANRYL 360

QY 361 AMKEDGRLLASKCVTDCEFFERLESNNYNTYRSRKYTSWYVALKRTGYKLGSKTGPQ 420
Db 361 AMKEDGRLLASKCVTDCEFFERLESNNYNTYRSRKYTSWYVALKRTGYKLGSKTGPQ 420

QY 421 KAILFLPMSAASDELPLQVLTLPHPNLHGPEILDVPST 457
Db 421 KAILFLPMSAASDELPLQVLTLPHPNLHGPEILDVPST 457

RESULT 2
US-08-809-156B-5
; Sequence 5, Application US/08090156B
; Patent No. 6472204
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; APPLICANT: Uemori, Takashi
; APPLICANT: Ueno, Takashi
; APPLICANT: Koyama, No. 6472204auto
; APPLICANT: Hashino, Kimikazu
; APPLICANT: Kato, Ikunoshin
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
; TITLE OF INVENTION: CELLS WITH RETROVIRUS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/809,156B
; FILING DATE: 07-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03254

; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 294382/1995
; FILING DATE: 13-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 051847/1996
; FILING DATE: 08-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 977.6507P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 457 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-809-156B-5

Query Match          100.0%; Score 2398; DB 4; Length 457;
Best Local Similarity 100.0%; Pred. No. 1.2e-171;
Matches 457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTDLRFNIGPDMRVMTWAPPSSIDLTLNLFVRYSPVKNEEDVAELISPSDNAVLTNLL 60
Db 1 PTDLRFNIGPDMRVMTWAPPSSIDLTLNLFVRYSPVKNEEDVAELISPSDNAVLTNLL 60

QY 61 PGTEYVSVSSVYEQHESTPLRGQKTGLDSTPGIDFSDITANSFTVHWIAPRATITGYR 120
Db 61 PGTEYVSVSSVYEQHESTPLRGQKTGLDSTPGIDFSDITANSFTVHWIAPRATITGYR 120

QY 121 IRHHPHFSGRPREDRVPHSRNSITLNLTPGTEYVSVIVALNGREESPLLIGQOSTVSD 180
Db 121 IRHHPHFSGRPREDRVPHSRNSITLNLTPGTEYVSVIVALNGREESPLLIGQOSTVSD 180

QY 181 VPRDLEVVAAATPTSLISWDAPAVTVRYRITYTGEGNSPVQEFVPGSKSTATISGLK 240
Db 181 VPRDLEVVAAATPTSLISWDAPAVTVRYRITYTGEGNSPVQEFVPGSKSTATISGLK 240

QY 241 PGVDYTIITVAVTGRGDSPASSKPIISINRYRTEIDKPSMAAGSITTLPALPEDGGSGAFP 300
Db 241 PGVDYTIITVAVTGRGDSPASSKPIISINRYRTEIDKPSMAAGSITTLPALPEDGGSGAFP 300

QY 301 GHFKDPKRLYCKNGGFFLRIHPDGRVDGVREKSDPHIKLOQAEERGVSIVKGVCANRYL 360
Db 301 GHFKDPKRLYCKNGGFFLRIHPDGRVDGVREKSDPHIKLOQAEERGVSIVKGVCANRYL 360

QY 361 AMKEDGRLLASKCVTDCEFFERLESNNYNTYRSRKYTSWYVALKRTGYKLGSKTGPQ 420
Db 361 AMKEDGRLLASKCVTDCEFFERLESNNYNTYRSRKYTSWYVALKRTGYKLGSKTGPQ 420

QY 421 KAILFLPMSAASDELPLQVLTLPHPNLHGPEILDVPST 457
Db 421 KAILFLPMSAASDELPLQVLTLPHPNLHGPEILDVPST 457

RESULT 3
US-07-959-369-8
; Sequence 8, Application US/07959369
; Patent No. 5302701
; GENERAL INFORMATION:
; APPLICANT: Hidetaka HASHI et al.
; TITLE OF INVENTION: No. 5302701el Functional Polypeptide
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
```



```

; LENGTH: 432 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-836-854-20

Query Match          94.1%; Score 2257; DB 2; Length 432;
Best Local Similarity 99.8%; Pred. No. 3.9e-161;
Matches 431; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTLRFTNIGDPTMRVTWAPPSSIDLTNFLVRYSPVKNEEDVAELSISSDNAVLTNLL 60
Db 1 PTLRFTNIGDPTMRVTWAPPSSIDLTNFLVRYSPVKNEEDVAELSISSDNAVLTNLL 60
QY 61 PGTEYVVSVSVEQHESTPLRGOKTGLDPTGIDFSDITANSFTVHMIAPRATITGYR 120
Db 61 PGTEYVVSVSVEQHESTPLRGOKTGLDPTGIDFSDITANSFTVHMIAPRATITGYR 120
QY 121 IRHHPHFSGRPREDRVPHSRNSITLNLTPGTEYVVSIVALNGREESPLLIGQQSTVSD 180
Db 121 IRHHPHFSGRPREDRVPHSRNSITLNLTPGTEYVVSIVALNGREESPLLIGQQSTVSD 180
QY 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVTPGSKSTATISGLK 240
Db 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVTPGSKSTATISGLK 240
QY 241 PGVDYITVYAVTGRGDSPASSKPIISINRYTEIDKPSMAAGSITTLPALPEDGGSGAPP 300
Db 241 PGVDYITVYAVTGRGDSPASSKPIISINRYTEIDKPSMAAGSITTLPALPEDGGSGAPP 300
QY 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLQQAEBRGVVSIKGVCANRYL 360
Db 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLQQAEBRGVVSIKGVCANRYL 360
QY 361 AKMEDGRLLASKCVTDECFFERLESNNYNTYRSKYSTWYVALKRTGYKLGSKTGPQ 420
Db 361 AKMEDGRLLASKCVTDECFFERLESNNYNTYRSKYSTWYVALKRTGYKLGSKTGPQ 420
QY 421 KAILFLPMSAAS 432
Db 421 KAILFLPMSAKS 432
```

RESULT 5

```

US-09-366-009-4
; Sequence 4, Application US/09366009
; Patent No. 6426042
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; Uemori, Takashi
; Koyama, No. 6426042uto
; Hashino, Kimikazu
; Kato, Ikunoshin
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
; CELLS WITH RETROVIRUS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/366,009
; FILING DATE: 02-Aug-1999
```

```

; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/809,156
; FILING DATE: <Unknown>
; APPLICATION NUMBER: JP 294382/1995
; FILING DATE: 13-NOV-1995
; APPLICATION NUMBER: JP 051847/1996
; FILING DATE: 08-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 977.6507P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 432 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-366-009-4

Query Match          94.1%; Score 2257; DB 4; Length 432;
Best Local Similarity 99.8%; Pred. No. 3.9e-161;
Matches 431; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTDLFTNIGDPTMRVTWAPPSSIDLTNFLVRYSPVKNEEDVAELSISSDNAVLTNLL 60
Db 1 PTDLFTNIGDPTMRVTWAPPSSIDLTNFLVRYSPVKNEEDVAELSISSDNAVLTNLL 60
QY 61 PGTEYVVSVSVEQHESTPLRGOKTGLDPTGIDFSDITANSFTVHMIAPRATITGYR 120
Db 61 PGTEYVVSVSVEQHESTPLRGOKTGLDPTGIDFSDITANSFTVHMIAPRATITGYR 120
QY 121 IRHHPHFSGRPREDRVPHSRNSITLNLTPGTEYVVSIVALNGREESPLLIGQQSTVSD 180
Db 121 IRHHPHFSGRPREDRVPHSRNSITLNLTPGTEYVVSIVALNGREESPLLIGQQSTVSD 180
QY 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVTPGSKSTATISGLK 240
Db 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVTPGSKSTATISGLK 240
QY 241 PGVDYITVYAVTGRGDSPASSKPIISINRYTEIDKPSMAAGSITTLPALPEDGGSGAPP 300
Db 241 PGVDYITVYAVTGRGDSPASSKPIISINRYTEIDKPSMAAGSITTLPALPEDGGSGAPP 300
QY 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLQQAEBRGVVSIKGVCANRYL 360
Db 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLQQAEBRGVVSIKGVCANRYL 360
QY 361 AKMEDGRLLASKCVTDECFFERLESNNYNTYRSKYSTWYVALKRTGYKLGSKTGPQ 420
Db 361 AKMEDGRLLASKCVTDECFFERLESNNYNTYRSKYSTWYVALKRTGYKLGSKTGPQ 420
QY 421 KAILFLPMSAAS 432
Db 421 KAILFLPMSAKS 432
```

RESULT 6

```

US-08-809-156B-4
; Sequence 4, Application US/08809156B
; Patent No. 6472204
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; Uemori, Takashi
; Koyama, No. 6472204uto
; Hashino, Kimikazu
; Kato, Ikunoshin
```

;; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET  
;; TITLE OF INVENTION: CELLS WITH RETROVIRUS  
;; NUMBER OF SEQUENCES: 39  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: WEISER & ASSOCIATES  
;; STREET: 230 South Fifteenth Street, Suite 500  
;; CITY: Philadelphia  
;; STATE: PA  
;; COUNTRY: USA  
;; ZIP: 19102  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent in Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/809,156B  
;; FILING DATE: 07-MAR-1997  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/JP96/03254  
;; FILING DATE: 07-NOV-1996  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 294382/1995  
;; FILING DATE: 13-NOV-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 051847/1996  
;; FILING DATE: 08-MAR-1996  
;; TELECOMMUNICATION INFORMATION:  
;; NAME: Weiser, Gerard J.  
;; REGISTRATION NUMBER: 19,763  
;; REFERENCE/DOCKET NUMBER: 977.6507P  
;; TELEPHONE: 215-875-8383  
;; TELEFAX: 215-875-8394  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 432 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; US-08-809-156B-4

Query Match 94.1%; Score 2257; DB 4; Length 432;  
Best Local Similarity 99.8%; Pred. No. 3.9e-161;  
Matches 431; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 PTDLFTNIGPDMRVMTWAPPIDLTNFLVRYSPVKNEDVAELSIKSPSDNAVLTNLL 60  
DB 1 PTDLFTNIGPDMRVMTWAPPIDLTNFLVRYSPVKNEDVAELSIKSPSDNAVLTNLL 60  
QY 61 PGTEYVSVSSVEQHESTPLRGOKTGLDPSPTGIDFSDITANSFTVHMIAPATITGYR 120  
DB 61 PGTEYVSVSSVEQHESTPLRGOKTGLDPSPTGIDFSDITANSFTVHMIAPATITGYR 120  
QY 121 IRHHPHFSGRPREDRVPHSRNITLNTLPGTEYVSVIVALNGREESPLLIQQQSTVSD 180  
DB 121 IRHHPHFSGRPREDRVPHSRNITLNTLPGTEYVSVIVALNGREESPLLIQQQSTVSD 180  
QY 181 VPRDLVVAATPTSLISWDAPAVTVRYRITYTGETGNSPVQEFVPGSKSTATISGLK 240  
DB 181 VPRDLVVAATPTSLISWDAPAVTVRYRITYTGETGNSPVQEFVPGSKSTATISGLK 240  
QY 241 PGVDYITVYAVTGRGDSPASSKPIINRYRTEIDKPSMAAGSITTLPALPEDGSGAPFP 300  
DB 241 PGVDYITVYAVTGRGDSPASSKPIINRYRTEIDKPSMAAGSITTLPALPEDGSGAPFP 300  
QY 301 GHFKDKRLKCKNGGFFLRHPDGRVDGVRKSDPHIKLQQAERGVSIVKGVCANRYL 360  
DB 301 GHFKDKRLKCKNGGFFLRHPDGRVDGVRKSDPHIKLQQAERGVSIVKGVCANRYL 360  
QY 361 AMKEDGRLLASKCVTDECFFERLESNNYNTYRSRKYTSWYVALKRTGYKLGSKTGPQ 420

DB 361 AMKEDGRLLASKCVTDECFFERLESNNYNTYRSRKYTSWYVALKRTGYKLGSKTGPQ 420  
QY 421 KAILFLPMSAAS 432  
DB 421 KAILFLPMSAKS 432  
RESULT 7  
US-07-959-369-9  
; Sequence 9, Application US/07959369  
; Patent No. 5302701  
; GENERAL INFORMATION:  
; APPLICANT: Hidetaka HASHI et al.  
; TITLE OF INVENTION: No. 5302701el Functional Polypeptide  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wenderoth, Lind & Ponack  
; STREET: 805 Fifteenth Street, N.W., #700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/959,369  
; FILING DATE: 19921013  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warren M. Cheek, Jr.  
; REGISTRATION NUMBER: 33,367  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-8850  
; TELEFAX:  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 432 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: polypeptide  
; HYPOTHETICAL:  
; ANTI-SENSE:  
; FRAGMENT TYPE:  
; ORIGINAL SOURCE:  
; ORGANISM:  
; STRAIN:  
; INDIVIDUAL ISOLATE:  
; DEVELOPMENTAL STAGE:  
; HAPLOTYPE:  
; TISSUE TYPE:  
; CELL TYPE:  
; CELL LINE:  
; ORGANELLE:  
; IMMEDIATE SOURCE:  
; LIBRARY:  
; CLONE:  
; POSITION IN GENOME:  
; CHROMOSOME/SEGMENT:  
; MAP POSITION:  
; UNITS:  
; FEATURE:  
; NAME/KEY:  
; LOCATION:

IDENTIFICATION METHOD:  
OTHER INFORMATION:  
PUBLICATION INFORMATION:  
AUTHORS:  
TITLE:  
JOURNAL:  
VOLUME:  
ISSUE:  
PAGES:  
DATE:  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO:  
US-07-959-369-9

Query Match 94.0%; Score 2254; DB 1; Length 432;  
Best Local Similarity 99.5%; Pred. No. 6.6e-161;  
Matches 430; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 PTDLFTNIGPDMRVMTWAPPSSIDLTFNLVRYSPVKNEEDVAELISGPSNNAVLTNLL 60  
Db 1 PTDLFTNIGPDMRVMTWAPPSSIDLTFNLVRYSPVKNEEDVAELISGPSNNAVLTNLL 60  
QY 61 PGTYYVSVSSVYEQHESTPLRGKQKTGLDSTGIDFSDITANSFTVHVIAPRATITGYR 120  
Db 61 PGTYYVSVSSVYEQHESTPLRGKQKTGLDSTGIDFSDITANSFTVHVIAPRATITGYR 120  
QY 121 IRHHPHFSGRPREDRVPHSRNSITLNTLTPGTYYVSVIVALNGREESPLLIGQOSTVSD 180  
Db 121 IRHHPHFSGRPREDRVPHSRNSITLNTLTPGTYYVSVIVALNGREESPLLIGQOSTVSD 180  
QY 181 VPRDLEWVAATPTSLISWDAPAVTVRYRYITYGETGNSPVQEFVPGSKSTATISGLK 240  
Db 181 VPRDLEWVAATPTSLISWDAPAVTVRYRYITYGETGNSPVQEFVPGSKSTATISGLK 240  
QY 241 PGVDYITVYAVTGRGDSPASSKPIISINRYRTEIDKPSMAAGSITTLPALPEDGGSGAPP 300  
Db 241 PGVDYITVYAVTGRGDSPASSKPIISINRYRTEIDKPSMAAGSITTLPALPEDGGSGAPP 300  
QY 301 GHFKDKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLOQAEERGVSIVKGVCANRYL 360  
Db 301 GHFKDKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLOQAEERGVSIVKGVCANRYL 360  
QY 361 AKMEDGRLASKVTDCECFEERLESNNYNTYRSRYTSWYVALKRTGYKLGSKTGPQ 420  
Db 361 AKMEDGRLASKVTDCECFEERLESNNYNTYRSRYTSWYVALKRTGYKLGSKTGPQ 420  
QY 421 KAILFLPMSAAS 432  
Db 421 KAILFLPMSAKS 432

RESULT 8  
US-08-836-854-21  
Sequence 21, Application US/08836854  
Patent No. 5824547  
GENERAL INFORMATION:  
APPLICANT: HASHINO, Kimikazu  
APPLICANT: MATSUSHITA, Hideyuki  
APPLICANT: KATO, Ikunoshin  
TITLE OF INVENTION: METHOD OF PRODUCTION OF TRANSPECTED CELLS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street N.W. Ste. 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/836.854  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP95/02425  
FILING DATE: 29-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 317721/1994  
FILING DATE: 29-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Browdy, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: HASHINO=1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-5197  
TELEFAX: (202) 737-3528  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 574 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-836-854-21  
Query Match 64.1%; Score 1536.5; DB 2; Length 574;  
Best Local Similarity 58.0%; Pred. No. 4.8e-107;  
Matches 344; Conservative 26; Mismatches 68; Indels 155; Gaps 14;  
QY 1 PTDLFTNIGPDMRVMTWAPPSSIDLTFNLVRYSPVKNEEDVAELISGPSNNAVLTNLL 60  
Db 1 PTDLFTNIGPDMRVMTWAPPSSIDLTFNLVRYSPVKNEEDVAELISGPSNNAVLTNLL 60  
QY 61 PGTYYVSVSSVYEQHESTPLRGKQKTGLDSTGIDFSDITANSFTVHVIAPRATITGYR 120  
Db 61 PGTYYVSVSSVYEQHESTPLRGKQKTGLDSTGIDFSDITANSFTVHVIAPRATITGYR 120  
QY 121 IRHHPHFSGRPREDRVPHSRNSITLNTLTPGTYYVSVIVALNGREESPLLIGQOSTVSD 180  
Db 121 IRHHPHFSGRPREDRVPHSRNSITLNTLTPGTYYVSVIVALNGREESPLLIGQOSTVSD 180  
QY 181 VPRDLEWVAATPTSLISWDAPAVTVRYRYITYGETGNSPVQEFVPGSKSTATISGLK 240  
Db 181 VPRDLEWVAATPTSLISWDAPAVTVRYRYITYGETGNSPVQEFVPGSKSTATISGLK 240  
QY 241 PGVDYITVYAVTGRGDSPASSKPIISINRYRTEIDKPSMAAGSITTL---PALPEDGGSGA 297  
Db 241 PGVDYITVYAVTGRGDSPASSKPIISINRYRTEIDKPSMAIPAPTDLKTQVTPSLSAQW 300  
QY 298 PPPGHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLOQAEERGVSIVKGVCAN 357  
Db 301 TPPN-----VLTGYRVVTP-----KEKTGPMKEINLAPDSSVVVSGLMVAT 344  
QY 358 RY---LAMKEDGRLASK-----CVTDEC----- 378  
Db 345 KYEVSVYALKD---TLTSRPAQGVVTTLENVSPRRARVTDATETITISWRTKTETITG 401  
QY 379 PFERLESNNYNTYRS-----RKYT----- 398  
Db 402 FQDVAVPANGQPTIORTIKPDVRSYITGLQPGTDYKIVLYTLNDNARSSPVVIDASTAI 461  
QY 399 -----SWYVALKRTGY-----KLASK-----TG- 417  
Db 462 DAPSNLRFATPNSLLVSWQPPRARITGYIIKYEKPGSPPREVVRPRPGVTEATITGL 521  
QY 418 -PGKAILFL-----PM--SAASDELUPQVTLPHNPHLHGPPEILDVPST 457  
Db 522 EPGTEVTIVIALKNNQKSEPLIGRKTDELQVTLPHNPHLHGPPEILDVPST 574



## RESULT 9

US-09-463-296-1  
; Sequence 1, Application US/09463296  
; Patent No. 6287864  
; GENERAL INFORMATION:  
; APPLICANT: BAGNIS, Claude  
; APPLICANT: IMBERT, Anne-Marie  
; APPLICANT: MANNONI, Patrice  
; TITLE OF INVENTION: GENE TRANSFER METHOD WITH THE USE OF SERUM-FREE MEDIUM  
; FILE REFERENCE: BAGNIS=1  
; CURRENT APPLICATION NUMBER: US/09/463,296  
; CURRENT FILING DATE: 2000-01-24  
; EARLIER APPLICATION NUMBER: PCT/JP98/03173  
; EARLIER FILING DATE: 1998-07-15  
; EARLIER APPLICATION NUMBER: JP 196772/1997  
; EARLIER FILING DATE: 1997-07-23  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 574  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-463-296-1

Query Match 64.1%; Score 1536.5; DB 3; Length 574;  
Best Local Similarity 58.0%; Pred. No. 4.8e-107;  
Matches 344; Conservative 26; Mismatches 68; Indels 155; Gaps 14;  
QY 1 PTLRFTNIGPDTRMTWAPPSPSIDLTNLFVRYSPVKNEEDVAELSISSPSDNAVLTNLL 60  
Db 1 PTLRFTNIGPDTRMTWAPPSPSIDLTNLFVRYSPVKNEEDVAELSISSPSDNAVLTNLL 60  
QY 61 PGTEYVSVSSVYEOHESTPLRGKQKTGLDPSGTIDFSDITANSFTVHMIAPRATTIGYR 120  
Db 61 PGTEYVSVSSVYEOHESTPLRGKQKTGLDPSGTIDFSDITANSFTVHMIAPRATTIGYR 120  
QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGREESPLLICQOSTVSD 180  
Db 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGREESPLLICQOSTVSD 180  
QY 181 VPRDLEVAATPTSLLSISWDAPAVTVRYRYITGETGNSPVQEFVPGSKSTATISGLK 240  
Db 181 VPRDLEVAATPTSLLSISWDAPAVTVRYRYITGETGNSPVQEFVPGSKSTATISGLK 240  
QY 241 PGVDYTYTVAVTGRGDSPASSKPIISINRYRTEIDKPSMAAGSITTL---PALPEDGSGA 297  
Db 241 PGVDYTYTVAVTGRGDSPASSKPIISINRYRTEIDKPSMAAGSITTL---PALPEDGSGA 297  
QY 298 FPGHFKDKRLCKNGGFFLRHPDGRVGVREKSDPHIKLOQAEERGVVSIKGVCAN 357  
Db 301 TPNP-----VLTGYRVVTP-----KEKTGPMKEINLAPDSSSVVVGMLVAT 344  
QY 358 RY---LAKMEDGRLASK-----CVTDEC----- 378  
Db 345 KYEVSVALKD---TLTSRPAQGVVTTLENVSPRRARVTDATETITISWRTKTETITG 401  
QY 379 FFERLEBSNNYTVRS-----RKYT----- 398  
Db 402 FOVDAPVANGQTPQRIKDPVRSYITITGLPGTDYKLYLTNDNARSPVVIDASTAI 461  
QY 399 -----SNVALKRTGY-----KLGSK-----TG- 417  
Db 462 DAPSNLRLFLATTNSLLVSPPRARITGVIIKYPGSPPEVVRPRRGVTEATITGL 521  
QY 418 -PGOKAILFL-----PM--SAASDELQPLVTLPHNLGCPETLDVPST 457  
Db 522 EPQTEYTYIVALKNNQKSEBELGRKKTDELQPLVTLPHNLGCPETLDVPST 574

## RESULT 10

US-09-366-009-24  
; Sequence 24, Application US/09366009  
; Patent No. 6426042

GENERAL INFORMATION:  
APPLICANT: Asada, Kiyozo  
Uemori, Takashi  
Ueno, Takashi  
Koyama, No. 6426042uto  
Hashino, Kimikazu  
Kato, Ikunoshin  
TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET  
CELLS WITH RETROVIRUS  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: WEISER & ASSOCIATES  
STREET: 230 South Fifteenth Street, Suite 500  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/366,009  
FILING DATE: 02-Aug-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/809,156  
FILING DATE: <Unknown>  
APPLICATION NUMBER: JP 294382/1995  
FILING DATE: 13-NOV-1995  
APPLICATION NUMBER: JP 051847/1996  
FILING DATE: 08-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Weiser, Gerard J.  
REGISTRATION NUMBER: 19,763  
REFERENCE/DOCKET NUMBER: 977.6507P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-875-8383  
TELEFAX: 215-875-8394  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 574 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 24:  
US-09-366-009-24

Query Match 64.1%; Score 1536.5; DB 4; Length 574;  
Best Local Similarity 58.0%; Pred. No. 4.8e-107;  
Matches 344; Conservative 26; Mismatches 68; Indels 155; Gaps 14;  
QY 1 PTLRFTNIGPDTRMTWAPPSPSIDLTNLFVRYSPVKNEEDVAELSISSPSDNAVLTNLL 60  
Db 1 PTLRFTNIGPDTRMTWAPPSPSIDLTNLFVRYSPVKNEEDVAELSISSPSDNAVLTNLL 60  
QY 61 PGTEYVSVSSVYEOHESTPLRGKQKTGLDPSGTIDFSDITANSFTVHMIAPRATTIGYR 120  
Db 61 PGTEYVSVSSVYEOHESTPLRGKQKTGLDPSGTIDFSDITANSFTVHMIAPRATTIGYR 120  
QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGREESPLLICQOSTVSD 180  
Db 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVIVALNGREESPLLICQOSTVSD 180  
QY 181 VPRDLEVAATPTSLLSISWDAPAVTVRYRYITGETGNSPVQEFVPGSKSTATISGLK 240  
Db 181 VPRDLEVAATPTSLLSISWDAPAVTVRYRYITGETGNSPVQEFVPGSKSTATISGLK 240  
QY 241 PGVDYTYTVAVTGRGDSPASSKPIISINRYRTEIDKPSMAAGSITTL---PALPEDGSGA 297  
Db 241 PGVDYTYTVAVTGRGDSPASSKPIISINRYRTEIDKPSMAAGSITTL---PALPEDGSGA 297



;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; FILING APPLICATION NUMBER: US/09/366,009  
;; FILING DATE: 02-Aug-1999  
;; CLASSIFICATION: <Unknown>  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/809,156  
;; FILING DATE: <Unknown>  
;; APPLICATION NUMBER: JP 294382/1995  
;; FILING DATE: 13-NOV-1995  
;; APPLICATION NUMBER: JP 051847/1996  
;; FILING DATE: 08-MAR-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Weiser, Gerard J.  
;; REGISTRATION NUMBER: 19,763  
;; REFERENCE/DOCKET NUMBER: 977.6507P  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 215-875-8394  
;; TELEFAX: 215-875-8394  
;; INFORMATION FOR SEQ ID NO: 8:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 489 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: <Unknown>  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
US-09-366-009-8

Query Match 63.7%; Score 1527; DB 4; Length 489;  
Best Local Similarity 65.5%; Pred. No. 2e-106;  
Matches 342; Conservative 7; Mismatches 75; Indels 98; Gaps 11;  
QY 1 PTDLRFNIGDTRMTVWAPPSPIDLTNLFVSPVKNEDVAELSPSPDNVAVLTNLL 60  
DB 1 PTDLRFNIGDTRMTVWAPPSPIDLTNLFVSPVKNEDVAELSPSPDNVAVLTNLL 60  
QY 61 PGTEYVSVSSVYEQHESTPLRGRQKTGLDPTGIDFSDITANSFTVHMIAPRATITGYR 120  
DB 61 PGTEYVSVSSVYEQHESTPLRGRQKTGLDPTGIDFSDITANSFTVHMIAPRATITGYR 120  
QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVISVALNGREESPLLIGQQSTVSD 180  
DB 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVISVALNGREESPLLIGQQSTVSD 180  
QY 181 VPRDLEVVAAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVTPGSKSTATISGLK 240  
DB 181 VPRDLEVVAAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVTPGSKSTATISGLK 240  
QY 241 PGVDYTTVYAVTGRGDSPASSKPIISINRYTEIDKPSMA-AGSITTLPALPEDGGSG--- 296  
DB 241 PGVDYTTVYAVTGRGDSPASSKPIISINRYTEIDKPSMGIRGLKGTGKEGDEGDFGFKG 300  
QY 297 -----APPPGHF-----KDPKRLYCKNG-----GFFLRIH 321  
DB 301 DMGIKGRGEITGPPGRGEDGEPGKGRGPNPGLPGLPPEKGLGVLGPGYGRQG 360  
QY 322 PDGRVD-----GVREKSDPHIKLQIQAEEGVVSIKGVCANRYLAMKE--- 364  
DB 361 PKGSIGFPFGANGEGKGRGTGPKGR-----CQRGTPGRGERGRGTGPKGPK 413  
QY 365 -----DGRLLASKVTDCEFFERLESNNYTSRKYTSWYVALKRTQYKLGSKTGP 419  
DB 414 GNSGGDGPAGPPG-----ERGPNGPQG-----TG--FPGKGPFG 447  
QY 420 QKAILFLP-----MSAADELQVLVTLPHPNLHGEILDVPEST 457  
DB 448 PPKDGLFPGHGGQASDELQVLVTLPHPNLHGEILDVPEST 489

RESULT 13  
US-08-809-156B-8

;; Sequence 8, Application US/08809156B  
;; Patent No. 6472204  
;; GENERAL INFORMATION:  
;; APPLICANT: Asada, Kiyozo  
;; APPLICANT: Umori, Takashi  
;; APPLICANT: Ueno, Takashi  
;; APPLICANT: Koyama, No. 6472204uto  
;; APPLICANT: Hashino, Kimikazu  
;; APPLICANT: Kato, Ikunoshin  
;; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET  
;; TITLE OF INVENTION: CELLS WITH RETROVIRUS  
;; NUMBER OF SEQUENCES: 39  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: WEISER & ASSOCIATES  
;; STREET: 230 South Fifteenth Street, Suite 500  
;; CITY: Philadelphia  
;; STATE: PA  
;; COUNTRY: USA  
;; ZIP: 19102  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/809,156B  
;; FILING DATE: 07-MAR-1997  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/JP96/03254  
;; FILING DATE: 07-NOV-1996  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 294382/1995  
;; FILING DATE: 13-NOV-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 051847/1996  
;; FILING DATE: 08-MAR-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Weiser, Gerard J.  
;; REGISTRATION NUMBER: 19,763  
;; REFERENCE/DOCKET NUMBER: 977.6507P  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 215-875-8394  
;; TELEFAX: 215-875-8394  
;; INFORMATION FOR SEQ ID NO: 8:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 489 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
US-08-809-156B-8

Query Match 63.7%; Score 1527; DB 4; Length 489;  
Best Local Similarity 65.5%; Pred. No. 2e-106;  
Matches 342; Conservative 7; Mismatches 75; Indels 98; Gaps 11;  
QY 1 PTDLRFNIGDTRMTVWAPPSPIDLTNLFVSPVKNEDVAELSPSPDNVAVLTNLL 60  
DB 1 PTDLRFNIGDTRMTVWAPPSPIDLTNLFVSPVKNEDVAELSPSPDNVAVLTNLL 60  
QY 61 PGTEYVSVSSVYEQHESTPLRGRQKTGLDPTGIDFSDITANSFTVHMIAPRATITGYR 120  
DB 61 PGTEYVSVSSVYEQHESTPLRGRQKTGLDPTGIDFSDITANSFTVHMIAPRATITGYR 120  
QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVISVALNGREESPLLIGQQSTVSD 180  
DB 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVISVALNGREESPLLIGQQSTVSD 180  
QY 181 VPRDLEVVAAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVTPGSKSTATISGLK 240  
DB 181 VPRDLEVVAAATPTSLISWDAPAVTVRYRITYGETGNSPVQEFVTPGSKSTATISGLK 240

QY 241 PGVDYITIVYAVTGRGDSPASSKPIISINRYTEIDKPSMA-AGSITTLPALPEDGGSG--- 296  
Db 241 PGVDYITIVYAVTGRGDSPASSKPIISINRYTEIDKPSMGIRGLKGTGKEKGEDGPFPGFKG 300  
QY 297 -----AFPPGHE-----KDPKLYCKNG-----GFFLRH 321  
Db 301 DMGIKDRGIGPPRGEDGPKRGKRGPNPDGPLGPPGKGLGVPLGPGRGQ 360  
QY 322 PDGRVD-----GVREKSDPHIKLOQAEERGVSIGVCANRYLAMKE--- 364  
Db 361 PKSGIGFPFGANGKGGRTGPKCPR-----GQRGPTGRGERGPRGITGPKGPK 413  
QY 365 -----DORLLASKVTCDECFERLESNNVNTYRSKYTSWYVALKRTGQYKLGSKTGP 419  
Db 414 GNSGDDGAGPPG-----BRGPNQPG-----TG--FGPKGPPG 447  
QY 420 QKAILFLP-----NSAASDELQVLTLPHPNLHGPEILDVPST 457  
Db 448 PPKDGLGPHGPGORGASDELQVLTLPHPNLHGPEILDVPST 489

RESULT 14

US-08-153-799-16  
; Sequence 16, Application US/08153799  
; Patent No. 5766883  
; GENERAL INFORMATION:  
; APPLICANT: Ballance, David J  
; APPLICANT: Goodey, Andrew R  
; TITLE OF INVENTION: Polypeptides  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: R Hain Swope, BOC Health Care Inc  
; STREET: 100 Mountain Avenue  
; CITY: Murray Hill  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07974  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/153,799  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/847975  
; FILING DATE: 06-MAR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 8909916.2  
; FILING DATE: 29-APR-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB90/00650  
; FILING DATE: 26-APR-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/775952  
; FILING DATE: 29-OCT-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Swope, R Hain  
; REGISTRATION NUMBER: 24864  
; REFERENCE/DOCKET NUMBER: 92H832  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (908) 665 2400  
; TELEFAX: (908) 771 6159  
; TELEX: 219484  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2231 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO

; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: Protein  
; LOCATION: 1..2231  
; OTHER INFORMATION: /note= "Human fibronectin"  
US-08-153-799-16  
  
Query Match 62.5%; Score 1499; DB 1; Length 2231;  
Best Local Similarity 51.4%; Pred No. 1.8e-103;  
Matches 346; Conservative 31; Mismatches 64; Indels 232; Gaps 17;  
  
QY 1 PTDLFTNIGPTDMRTWAPPSIDLNTNPLVRYSPVKNEEDVAELISPSDNNAVLTNLL 60  
Db 1239 PTDLFTNIGPTDMRTWAPPSIDLNTNPLVRYSPVKNEEDVAELISPSDNNAVLTNLL 1298  
  
QY 61 PCTEYVSVSSVYEQHESTPLRGROKGTGLDSPTGTDGDFDITANSTVFWIAPRAITGYR 120  
Db 1299 PCTEYVSVSSVYEQHESTPLRGROKGTGLDSPTGTDGDFDITANSTVFWIAPRAITGYR 1358  
  
QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGCTEYVSVIVALNGREESPLLIQOOSTVSD 180  
Db 1359 IRHHPHFSGRPREDRVPHSRNSITLTNLTGCTEYVSVIVALNGREESPLLIQOOSTVSD 1418  
  
QY 181 VPRDLEVVAATPTSLISWDAPAVTVRYRITYGETGNSPVQVEFTVPGSKSTATISGLK 240  
Db 1419 VPRDLEVVAATPTSLISWDAPAVTVRYRITYGETGNSPVQVEFTVPGSKSTATISGLK 1478  
  
QY 241 PGVDYITIVYAVTGRGDSPASSKPIISINRYTEIDKPSMA-----AGSITTLPA--- 288  
Db 1479 PGVDYITIVYAVTGRGDSPASSKPIISINRYTEIDKPSMQVTDVQDNISVKWLPSSPV 1538  
  
QY 289 -----LPEDGGSGAFPPGHFKDPK-----RLYCKN----- 313  
Db 1539 TGYRVTTTPKNG-----PGTKTKTAGPDQTEMIEGLQPTVEYVSVYAQNPGESQP 1592  
  
QY 314 -----GGFFLRHPDGRVDGVREKSDPHI 337  
Db 1593 LVQTAVTITPAPTDLKFTQVTPTSLSAQWTPDVQLTGYRVVTP-----KQKTGPMK 1645  
  
QY 338 KLOQAERGVSIGVCANRY-----LAMKEDGRLLASK-----CVT 375  
Db 1646 EINLAPDSSSVVSGLMVATKYEVSVYALKD---TLTSRPAQGVVTTLENVSPRRARVT 1702  
  
QY 376 DEC-----FFERLESNNYTVRS-----RKYT----- 398  
Db 1703 DATETITISWTKTETITGQVDVAVPANGQTPQRTIKPDVRSYITIGLOPCTDYKIYL 1762  
  
QY 399 -----SWYVALKRITQY-----KLGSK 415  
Db 1763 YTLNDNARSSPVVIDASTAIDAPSNLRFLATTPNSLLVSWQPPRARITGYIKYKPGSP 1822  
  
QY 416 -----TG--PGQKAILFL-----PM--SAASDELQVLTLPHP 444  
Db 1823 PREVVPRPRPGVTEATITGLEPFGTEYTIYVIALKNNKSEFLGRKTKDELQVLTLPHP 1882  
  
QY 445 NLHGPEILDVPST 457  
Db 1883 NLHGPEILDVPST 1895

RESULT 15

US-08-836-854-5  
; Sequence 5, Application US/08836854  
; Patent No. 5824547  
; GENERAL INFORMATION:  
; APPLICANT: HASHINO, Kimikazu  
; APPLICANT: MATSUSHITA, Hideyuki  
; APPLICANT: KATO, Ikunoshin  
; TITLE OF INVENTION: METHOD OF PRODUCTION OF TRANSFECTED CELLS  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Browdy and Neimark

STREET: 419 Seventh Street N.W. Ste. 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/836,854  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP95/02425  
FILING DATE: 29-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 317721/1994  
FILING DATE: 29-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Browdy, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: HASHINO-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-5197  
TELEFAX: (202) 737-3528  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 302 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-836-854-5.

Query Match 61.9%; Score 1485.5; DB 2; Length 302;  
Best Local Similarity 66.1%; Pred. No. 1.4e-103;  
Matches 302; Conservative 0; Mismatches 0; Indels 155; Gaps 1;  
  
Qy 1 PTDLRFNIGPDTMRVTWAPPPSIDLTFNLVRYSPVKNEEDVAELSI<sup>60</sup>  
Db |||||||  
Qy 1 PTDLRFNIGPDTMRVTWAPPPSIDLTFNLVRYSPVKNEEDVAELSI<sup>60</sup>  
Db |||||||  
  
Qy 61 PGTEYVVS<sup>120</sup>  
Db |||||||  
  
Qy 61 PGTEYVVS<sup>120</sup>  
Db |||||||  
  
Qy 121 IRHHPHFGRPREDRVPHSRNSITLNLTPGTEYVVS<sup>180</sup>  
Db |||||||  
  
Qy 121 IRHHPHFGRPREDRVPHSRNSITLNLTPGTEYVVS<sup>180</sup>  
Db |||||||  
  
Qy 181 VPRDLEVAATPTSLLSWAPAVTVRYRITYGETGNSPVQEF<sup>240</sup>  
Db |||||||  
  
Qy 181 VPRDLEVAATPTSLLSWAPAVTVRYRITYGETGNSPVQEF<sup>240</sup>  
Db |||||||  
  
Qy 241 PGVDYTI<sup>300</sup>  
Db |||||||  
  
Qy 241 PGVDYTI<sup>276</sup>  
Db |||||||  
  
Qy 301 GHFKDPKRLYCKNGGFFLRHPDGRVDGVRKSDPHIKLQ<sup>360</sup>  
Db |||||||  
  
Qy 277 -----<sup>276</sup>  
Db |||||||  
  
Qy 361 AMKEDGRL<sup>420</sup>  
Db |||||||  
  
Qy 277 -----<sup>276</sup>  
Db |||||||  
  
Qy 421 KAILFLPMSAASDEL<sup>457</sup>  
Db |||||||  
  
Qy 277 -----<sup>302</sup>  
Db |||||||

Search completed: May 3, 2004, 13:12:34  
Job time : 25.4541 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 13:11:14 ; Search time 54.7263 Seconds  
(without alignments)

2314.713 Million cell updates/sec

Title: US-09-775-964-5

Perfect score: 2398

Sequence: 1 PDLRFTNIGPDMRTWAP.....LVTLPHPNLGHPILDVPST 457

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1138120 seqs, 277189581 residues

Total number of hits satisfying chosen parameters: 1138120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2398	100.0	457	10	US-09-775-964-5
2	2257	94.1	432	10	US-09-775-964-4
3	1536.5	64.1	574	10	US-09-775-964-24
4	1527	63.7	489	10	US-09-775-964-8
5	1485.5	61.9	302	10	US-09-775-964-29
6	1474	61.5	2320	12	US-10-236-392-2
7	1474	61.5	2320	14	US-10-279-733-8
8	1474	61.5	2328	12	US-10-182-936A-98
9	1474	61.5	2328	14	US-10-171-311-64
10	1474	61.5	2328	15	US-10-236-031B-70
11	1474	61.5	2328	15	US-10-374-979-98
12	1474	61.5	2355	15	US-10-144-194A-104
13	1474	61.5	2355	15	US-10-360-101-235
14	1474	61.5	2355	16	US-10-447-161-3
15	1474	61.5	2386	10	US-09-961-403-1

16	1463.5	61.0	472	10	US-09-775-964-21	Sequence 21, Appl
17	1463.5	61.0	549	10	US-09-775-964-23	Sequence 23, Appl
18	1454	60.6	826	10	US-09-775-964-14	Sequence 14, Appl
19	1448.5	60.4	464	10	US-09-775-964-7	Sequence 7, Appl
20	1442	60.1	457	10	US-09-775-964-22	Sequence 22, Appl
21	1437	59.9	599	14	US-10-279-733-9	Sequence 9, Appl
22	1437	59.9	2220	12	US-10-236-392-4	Sequence 4, Appl
23	1421	59.3	274	10	US-09-775-964-25	Sequence 25, Appl
24	840.5	35.1	501	9	US-09-934-706-4	Sequence 4, Appl
25	824.5	34.4	196	12	US-10-389-821-6	Sequence 6, Appl
26	824.5	34.4	210	9	US-09-902-773A-4	Sequence 8, Appl
27	824.5	34.4	210	12	US-10-372-653-8	Sequence 8, Appl
28	821	34.2	158	9	US-09-826-210-2	Sequence 2, Appl
29	820.5	34.2	386	12	US-10-344-634-14	Sequence 14, Appl
30	820	34.2	155	9	US-09-822-485-5	Sequence 5, Appl
31	820	34.2	155	9	US-09-802-365-8	Sequence 8, Appl
32	820	34.2	155	9	US-09-251-263-10	Sequence 10, Appl
33	820	34.2	155	9	US-09-425-021-10	Sequence 10, Appl
34	820	34.2	155	9	US-09-886-856-8	Sequence 8, Appl
35	820	34.2	155	9	US-09-749-728B-7	Sequence 7, Appl
36	820	34.2	155	10	US-09-902-460-2	Sequence 17, Appl
37	820	34.2	155	10	US-09-345-373-17	Sequence 3, Appl
38	820	34.2	155	10	US-09-775-964-3	Sequence 28, Appl
39	820	34.2	155	14	US-10-081-347-28	Sequence 17, Appl
40	820	34.2	155	14	US-10-075-446-17	Sequence 11, Appl
41	820	34.2	155	14	US-10-189-360-11	Sequence 14, Appl
42	820	34.2	155	14	US-10-192-988-14	Sequence 5, Appl
43	820	34.2	155	14	US-10-374-207-5	Sequence 5, Appl
44	820	34.2	155	14	US-10-395-541-5	Sequence 4, Appl
45	820	34.2	155	14	US-10-123-481-4	Sequence 4, Appl

#### ALIGNMENTS

#### RESULT 1

US-09-775-964-5  
; Sequence 5, Application US/09775964  
; Publication No. US20030087437A1  
; GENERAL INFORMATION:  
; APPLICANT: Asada, Kiyozo  
; Uemori, Takashi  
; Ueno, Takashi  
; Koyama, No. US20030087437A1uto  
; Hashino, Kimikazu  
; Kato, Ikunoshin  
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET  
; CELLS WITH RETROVIRUS  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: WEISER & ASSOCIATES  
; STREET: 230 South Fifteenth Street, Suite 500  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19102

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/775,964  
FILING DATE: 20-Feb-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/366,009  
FILING DATE: 02-Aug-1999  
APPLICATION NUMBER: 08/809,156  
FILING DATE: <Unknown>  
APPLICATION NUMBER: JP 294382/1995  
FILING DATE: 13-NOV-1995  
APPLICATION NUMBER: JP 051847/1996

```
/
/ FILING DATE: 08-MAR-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Weiser, Gerard J.
/ REGISTRATION NUMBER: 19,763
/ REFERENCE/DOCKET NUMBER: 977.6507P
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 215-875-8383
/ TELEFAX: 215-875-8394
/ INFORMATION FOR SEQ ID NO: 5:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 457 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: <Unknown>
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-775-964-5

Query Match 100.0%; Score 2398; DB 10; Length 457;
Best Local Similarity 100.0%; Pred. No. 3.7e-185;
Matches 457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTDLRTNIGPDMRTVWAPPSPIDLTNPLVRYSPVKNEEDVAELSIQSDNAVLTNLL 60
DB 1 PTDLRTNIGPDMRTVWAPPSPIDLTNPLVRYSPVKNEEDVAELSIQSDNAVLTNLL 60
QY 61 PGTEYVSVSSVYEOHSTPLRGROKTGLDSDPTGIDFSDITANSFTVHWIAPRAITGYR 120
DB 61 PGTEYVSVSSVYEOHSTPLRGROKTGLDSDPTGIDFSDITANSFTVHWIAPRAITGYR 120
QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVSIVALNGREESPLLIGQQSTVSD 180
DB 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVSIVALNGREESPLLIGQQSTVSD 180
QY 181 VPRDLVVAATPSSLISWDAPAVTVRYRITYTGTSNPSVQEFVPGSKSTATISGLK 240
DB 181 VPRDLVVAATPSSLISWDAPAVTVRYRITYTGTSNPSVQEFVPGSKSTATISGLK 240
QY 241 PGVDYTIIVYAVTGRGSDPASSKPIINVTREIDKPSMAAGSITTLPALPEDGGGAPPP 300
DB 241 PGVDYTIIVYAVTGRGSDPASSKPIINVTREIDKPSMAAGSITTLPALPEDGGGAPPP 300
QY 301 GHFKDPKRLCYCKNGGFLLRIHPDGRVDGVREKSDPHIKLQQAERGVVSIKGVCANRYL 360
DB 301 GHFKDPKRLCYCKNGGFLLRIHPDGRVDGVREKSDPHIKLQQAERGVVSIKGVCANRYL 360
QY 361 AMKEDGRLLASKCVTDCEFFERLESNNYNTYRSKYTSWYVALKRTGOYKLGSKTGPQ 420
DB 361 AMKEDGRLLASKCVTDCEFFERLESNNYNTYRSKYTSWYVALKRTGOYKLGSKTGPQ 420
QY 421 KAILFLPMSAASDELPLQVTLPHNPHGPEILDVPST 457
DB 421 KAILFLPMSAASDELPLQVTLPHNPHGPEILDVPST 457

RESULT 2
US-09-775-964-4
; Sequence 4, Application US/09775964
; Publication No. US20030087437A1
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; Umori, Takashi
; Ueno, Takashi
; Kovama, No. US20030087437A1uto
; Hashino, Kimikazu
; Kato, Ikunoshin
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
; CELLS WITH RETROVIRUS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
```

```
/
/ STATE: PA
/ COUNTRY: USA
/ ZIP: 19102
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/775,964
/ FILING DATE: 20-Feb-2001
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/09/366,009
/ FILING DATE: 02-Aug-1999
/ APPLICATION NUMBER: 08/809,156
/ FILING DATE: <Unknown>
/ APPLICATION NUMBER: JP 294382/1995
/ FILING DATE: 13-NOV-1995
/ APPLICATION NUMBER: JP 051847/1996
/ FILING DATE: 08-MAR-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Weiser, Gerard J.
/ REGISTRATION NUMBER: 19,763
/ REFERENCE/DOCKET NUMBER: 977.6507P
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 215-875-8383
/ TELEFAX: 215-875-8394
/ INFORMATION FOR SEQ ID NO: 4:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 432 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: <Unknown>
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-775-964-4

Query Match 94.1%; Score 2257; DB 10; Length 432;
Best Local Similarity 99.8%; Pred. No. 8.4e-174;
Matches 431; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTDLFTNIGPDMRTVWAPPSPIDLTNPLVRYSPVKNEEDVAELSIQSDNAVLTNLL 60
DB 1 PTDLFTNIGPDMRTVWAPPSPIDLTNPLVRYSPVKNEEDVAELSIQSDNAVLTNLL 60
QY 61 PGTEYVSVSSVYEOHSTPLRGROKTGLDSDPTGIDFSDITANSFTVHWIAPRAITGYR 120
DB 61 PGTEYVSVSSVYEOHSTPLRGROKTGLDSDPTGIDFSDITANSFTVHWIAPRAITGYR 120
QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVSIVALNGREESPLLIGQQSTVSD 180
DB 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVSVSIVALNGREESPLLIGQQSTVSD 180
QY 181 VPRDLVVAATPSSLISWDAPAVTVRYRITYTGTSNPSVQEFVPGSKSTATISGLK 240
DB 181 VPRDLVVAATPSSLISWDAPAVTVRYRITYTGTSNPSVQEFVPGSKSTATISGLK 240
QY 241 PGVDYTIIVYAVTGRGSDPASSKPIINVTREIDKPSMAAGSITTLPALPEDGGGAPPP 300
DB 241 PGVDYTIIVYAVTGRGSDPASSKPIINVTREIDKPSMAAGSITTLPALPEDGGGAPPP 300
QY 301 GHFKDPKRLCYCKNGGFLLRIHPDGRVDGVREKSDPHIKLQQAERGVVSIKGVCANRYL 360
DB 301 GHFKDPKRLCYCKNGGFLLRIHPDGRVDGVREKSDPHIKLQQAERGVVSIKGVCANRYL 360
QY 361 AMKEDGRLLASKCVTDCEFFERLESNNYNTYRSKYTSWYVALKRTGOYKLGSKTGPQ 420
DB 361 AMKEDGRLLASKCVTDCEFFERLESNNYNTYRSKYTSWYVALKRTGOYKLGSKTGPQ 420
QY 421 KAILFLPMSAAS 432
DB 421 KAILFLPMSAKS 432
```



```

RESULT 3
US-09-775-964-24
; Sequence 24, Application US/09775964
; Publication No. US20030087437A1
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; Uemori, Takashi
; Koyama, No. US20030087437A1uto
; Hashino, Kimikazu
; Kato, Ikunoshin
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
; CELLS WITH RETROVIRUS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/775,964
; FILING DATE: 20-Feb-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/366,009
; FILING DATE: 02-Aug-1999
; APPLICATION NUMBER: 08/809,156
; FILING DATE: <Unknown>
; APPLICATION NUMBER: JP 294382/1995
; FILING DATE: 13-NOV-1995
; APPLICATION NUMBER: JP 051847/1996
; FILING DATE: 08-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 977.6507P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 574 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 24:
US-09-775-964-24

Query Match 64.1%; Score 1536.5; DB 10; Length 574;
Best Local Similarity 58.0%; Pred. No. 2e-115;
Matches 344; Conservative 26; Mismatches 68; Indels 155; Gaps 14;

Qy 1 PTDLRFTNIGDPTMRVTWAPPSPIDLTNFLVRYSPVKNEEDVAELSTSPSDNAVLTNLL 60
Db 1 PTDLRFTNIGDPTMRVTWAPPSPIDLTNFLVRYSPVKNEEDVAELSTSPSDNAVLTNLL 60
Qy 61 PGTEYVSVSSVYQHESTPLRQKTKGLDPSPTGIDFSDITANSFTVHMIAPRATITGYR 120
Db 61 PGTEYVSVSSVYQHESTPLRQKTKGLDPSPTGIDFSDITANSFTVHMIAPRATITGYR 120
Qy 121 IRHHPHFSGRPREDRVPFHSRNITLTNLTGTEYVSVISVALNGREESPLLICQQTSD 180
Db 121 IRHHPHFSGRPREDRVPFHSRNITLTNLTGTEYVSVISVALNGREESPLLICQQTSD 180

181 VPRDLEVAATPTSLLSISWDAPAVTVRYRITYTGETGNSPVQEFVPGSKSTATISGLK 240
181 VPRDLEVAATPTSLLSISWDAPAVTVRYRITYTGETGNSPVQEFVPGSKSTATISGLK 240
241 PGVDYTIITYAVTGRGDSPASSKPI SINRYRTEIDKPSMAAGSITTL---PALPEDGSGA 297
241 PGVDYTIITYAVTGRGDSPASSKPI SINRYRTEIDKPSMAIPAPFTDLKFTQVTPPTSLSAQW 300
298 FPPGHFKDPKRLYCKNGGFFLRHDPGRVDGVRKSDPHIKLQAEERGVWSIKGVCAN 357
301 TPN-----VQLTGYRVVTP-----KEKTPMKEINLADSSSVVVSGLMVAT 344
358 RY-----LAKKEDGRLLASK-----CVTDEC----- 378
345 KYEVSVYALKD---TLTSRPAQGVTTLENVSPRRARVTDATETITISWRTKTETITG 401
379 FFERLESNNYTYRS-----RKYT----- 398
402 FQVDAVPANGQTPQRTIKPDRVRSYTTITGLQPGTDYKIYLYTLNDNARSGPVVIDASTAI 461
399 -----SWYVALKRTGOY-----KLGSK-----TG- 417
462 DAPSNURFLATTPNSLLVSWQPPRARITGYIIKYEKPGSPPREVVRPRRCVTEATITGL 521
418 -PQKAILFL-----PM--SAASDELQVLTLPHPNLHGPEILDVPST 457
522 EPQTEVTIVIALKNNKQKSEPLIGRKKTDLPQVLTLPHPNLHGPEILDVPST 574

RESULT 4
US-09-775-964-8
; Sequence 8, Application US/09775964
; Publication No. US20030087437A1
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; Uemori, Takashi
; Koyama, No. US20030087437A1uto
; Hashino, Kimikazu
; Kato, Ikunoshin
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
; CELLS WITH RETROVIRUS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/775,964
; FILING DATE: 20-Feb-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/366,009
; FILING DATE: 02-Aug-1999
; APPLICATION NUMBER: 08/809,156
; FILING DATE: <Unknown>
; APPLICATION NUMBER: JP 294382/1995
; FILING DATE: 13-NOV-1995
; APPLICATION NUMBER: JP 051847/1996
; FILING DATE: 08-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 977.6507P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 574 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 24:
US-09-775-964-24
```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 489 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-775-964-8

Query Match 63.7%; Score 1527; DB 10; Length 489;
Best Local Similarity 65.5%; Pred. No. 9.5e-115;
Matches 342; Conservative 7; Mismatches 75; Indels 98; Gaps 11;

QY 1 PTDLRFNIGPDMRVTWAPPSSIDLTNLFVRYSPVKNEEDVAELISPSDNAVLTNLL 60
DB 1 PTDLRFNIGPDMRVTWAPPSSIDLTNLFVRYSPVKNEEDVAELISPSDNAVLTNLL 60
QY 61 PGTEYVVSVSSVYEQHESPTLRGQKTGLDSDPTGIDFSDITANSFTVHWIAPRATITGYR 120
DB 61 PGTEYVVSVSSVYEQHESPTLRGQKTGLDSDPTGIDFSDITANSFTVHWIAPRATITGYR 120
QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVVSIVALNGREESPLLIGQQSTVSD 180
DB 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVVSIVALNGREESPLLIGQQSTVSD 180
QY 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYTGEGNSPVQEFVPGSKSTATISGLK 240
DB 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYTGEGNSPVQEFVPGSKSTATISGLK 240
QY 241 PGVDYTIITVAVTGRGSDPASSKPISINRYTEIDKPSMA-AGSITLTPALPEDGGSG--- 296
DB 241 PGVDYTIITVAVTGRGSDPASSKPISINRYTEIDKPSMA-AGSITLTPALPEDGGSG--- 296
QY 297 -----AFPPGHF-----KDKRLCYKNG-----GFLRIH 321
DB 301 DMGIKGRGIBGIPGPRGEDCGPKGRGPGNDGPGPLGPGKGKLGVLPGYPGRG 360
QY 322 PDGRVD-----GVREKSDPHIKLOQAEERGVSIVKVCANRYLAMKE---- 364
DB 361 PKGSI GPFPGGANGKGGRTGPKGPR-----GQRGTPGRGERGRTGPKGPK 413
QY 365 -----DGRLLASCVTDECFERLESNNYNTYRSRKYTSWYVALKRTQYKLGSKTGP 419
DB 414 GNSGGDGPAGPPG-----ERGNPGQP-----TG--FPGKGPFG 447
QY 420 QKAILFLP-----MSAASDELPLQVLTLPHPNLHGPEILDVPST 457
DB 448 PPKDGLPGHPGQRGASDELPLQVLTLPHPNLHGPEILDVPST 489

RESULT 5
US-09-775-964-29
; Sequence 29, Application US/09775964
; Publication No. US20030087437A1
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; Umori, Takashi
; Ueno, Takashi
; Koyama, No. US20030087437Aluto
; Hashino, Kimikazu
; Kato, Ikunoshin
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
; CELLUS WITH RETROVIRUS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA

```

```

; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/775,964
; FILING DATE: 20-Feb-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/366,009
; FILING DATE: 02-Aug-1999
; APPLICATION NUMBER: 08/809,156
; FILING DATE: <Unknown>
; APPLICATION NUMBER: JP 294382/1995
; FILING DATE: 13-NOV-1995
; APPLICATION NUMBER: JP 051847/1996
; FILING DATE: 08-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 977.6507P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 302 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 29:
US-09-775-964-29

Query Match 61.9%; Score 1485.5; DB 10; Length 302;
Best Local Similarity 66.1%; Pred. No. 1.1e-111;
Matches 302; Conservative 0; Mismatches 0; Indels 155; Gaps 1;

QY 1 PTDLFTNIGPDMRVTWAPPSSIDLTNLFVRYSPVKNEEDVAELISPSDNAVLTNLL 60
DB 1 PTDLFTNIGPDMRVTWAPPSSIDLTNLFVRYSPVKNEEDVAELISPSDNAVLTNLL 60
QY 61 PGTEYVVSVSSVYEQHESPTLRGQKTGLDSDPTGIDFSDITANSFTVHWIAPRATITGYR 120
DB 61 PGTEYVVSVSSVYEQHESPTLRGQKTGLDSDPTGIDFSDITANSFTVHWIAPRATITGYR 120
QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVVSIVALNGREESPLLIGQQSTVSD 180
DB 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVVSIVALNGREESPLLIGQQSTVSD 180
QY 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYTGEGNSPVQEFVPGSKSTATISGLK 240
DB 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYTGEGNSPVQEFVPGSKSTATISGLK 240
QY 241 PGVDYTIITVAVTGRGSDPASSKPISINRYTEIDKPSMAAGSIITLTPALPEDGGSGAPPP 300
DB 241 PGVDYTIITVAVTGRGSDPASSKPISINRYTEIDKP----- 276
QY 301 GHFKDPKRLCYKNGGFFLRHDPGRVDGVRKSDPHIKLQLAERGVVSIVKVCANRYL 360
DB 277 ----- 276
QY 361 AWKEDGRLLASKCVTDECFERLESNNYNTYRSRKYTSWYVALKRTQYKLGSKTGPQG 420
DB 277 ----- 276
QY 421 KAILFLPMSAASDELPLQVLTLPHPNLHGPEILDVPST 457
DB 277 -----SDELPLQVLTLPHPNLHGPEILDVPST 302

```

```
RESULT 6
US-10-236-392-2
; Sequence 2, Application US/10236392
; Publication No. US20040067490A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David W
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Burgess, Catherine, E
; APPLICANT: Casman, Stacie J
; APPLICANT: Catterton, Elna
; APPLICANT: Chapoval, Andrei
; APPLICANT: Crabtree, Julie
; APPLICANT: Edinger, Shlomit, R
; APPLICANT: Ellerman, Karen
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gorman, Linda
; APPLICANT: Grosse, William M
; APPLICANT: Gusev, Vladamir
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Larocheille, William J
; APPLICANT: Li, Li
; APPLICANT: MacDougall, John R
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Miller, Charles E
; APPLICANT: Millet, Isabelle
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Patturajan, Meera
; APPLICANT: Pena, Carol A
; APPLICANT: Peyman, John A
; APPLICANT: Rastelli, Luca
; APPLICANT: Reiger, Daniel K
; APPLICANT: Rothenberg, Mark E
; APPLICANT: Shenoy, Suresh
; APPLICANT: Shimkets, Richard A
; APPLICANT: Smithson, Glennda
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442A
; CURRENT APPLICATION NUMBER: US/10/236,392
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: US09/540,763
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US60/390,155
; PRIOR FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: US09/635,949
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: US60/318,765
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US60/357,303
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: US60/367,753
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER: US60/369,479
; PRIOR FILING DATE: 2002-04-02
; PRIOR APPLICATION NUMBER: US09/659,634
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/318,130
; PRIOR FILING DATE: 2001-09-07
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 794
; SOFTWARE: Custom
; SEQ ID NO 2
; TYPE: PRT
; LENGTH: 2320
; ORGANISM: Homo sapiens
US-10-236-392-2
Query Match 61.5%; Score 1474; DB 12; Length 2320;
Best Local Similarity 46.8%; Pred. No. 1.6e-109;
Matches 351; Conservative 28; Mismatches 75; Indels 296; Gaps 17;

QY 1 PTDLRFNTNIGPDTRMTWAPPSIDLTNLFVRYSPVKNEEDVAELSISSPDNAVLTNLL 60
Db 1235 PTDLRFNTNIGPDTRMTWAPPSIDLTNLFVRYSPVKNEEDVAELSISSPDNAVLTNLL 1294
QY 61 PGTEYVVSUSSVYEOHSTPLRGROKTGLDSDPTGIDFSDITANSFTVHWTAPRATITGYR 120
Db 1295 PGTEYVVSUSSVYEOHSTPLRGROKTGLDSDPTGIDFSDITANSFTVHWTAPRATITGYR 1354
QY 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVVSIVALNGREESPLLIGQOSTVSD 180
Db 1355 IRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVVSIVALNGREESPLLIGQOSTVSD 1414
QY 181 VPRDLEVAATPTSLISWDAPAVTVYRITVGETGNSPVQEFVPGSKSTATISGLK 240
Db 1415 VPRDLEVAATPTSLISWDAPAVTVYRITVGETGNSPVQEFVPGSKSTATISGLK 1474
QY 241 PGVDYITVYAVTGRGSDSPASSKPI SINRYTRTEIDKPS-----M 278
Db 1475 PGVDYITVYAVTGRGSDSPASSKPI SINRYTRTEIDKPSOMQVTDVQDNSISVKWLPSSSPV 1534
QY 279 AAGSITTL----- 287
Db 1535 TGYRVTTTPKNGPGPTKKTAGDQDTEMIEGLQPTVEYVVSVAQNPSGESQPLVQTAV 1594
QY 288 -----ALPEDGGSGAFP-PGHFKDPKRL 309
Db 1595 TNIDRPKGLAFTDVDVDSIKIAWESPGQGVSRVRYTSSPEDGIHSLFPAPDGEEDTAEL 1654
QY 310 YCKNGG-----PFLRIH-----PDGRV 326
Db 1655 QGLRPGSEYTVSVVALHDDMESQPLIGTOSTAIPAPTDLKFTQVTPTSLSAOWTPPNVQL 1714
QY 327 DGVR-----EKSDPHIKLQAEERGVSIGVCANRY-----LAMKEDGRLLASK--- 372
Db 1715 TGYRVRVTPKEKTGPMKEINLAPDSSVVVSGLMWATKEVSVYALKD---TLTSRPAQG 1771
QY 373 -----CVTDEC-----FFFERLESNNYNTYRS-----R 395
Db 1772 VVTTLNENSPRRARVTDATETITISWRTKTETITGFQVDAVPANGQTPIQRTIKPDVR 1831
QY 396 KYT----- 403
Db 1832 SYTITGLQPGTDYKYLYTLNDNARSSPVVIDASTAIDAPSNLRFATTNPSLLVSWQPP 1891
QY 404 LKRTGQY-----KLGSK-----TG--PGQKAILFL-----PM- 428
Db 1892 RARITGYIIKYKPGSPPREVVPRPRPGVTEATITGLEFGTEYTIIVIALKNNQKSEPLI 1951
QY 429 -SAASDELQVLTLPHPNLHGPEILDVPST 457
Db 1952 GRKKTDDELQVLTLPHPNLHGPEILDVPST 1981

RESULT 7
US-10-279-733-8
; Sequence 8, Application US/10279733
; Publication No. US20030194400A1
; GENERAL INFORMATION:
; APPLICANT: The Scripps Research Institute
; APPLICANT: Liu, Cheng
; APPLICANT: Edgington, Thomas, S.
; TITLE OF INVENTION: Targeted Thrombosis
; FILE REFERENCE: 1361.016W01
; CURRENT APPLICATION NUMBER: US/10/279,733
; CURRENT FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: US 60/336331
; PRIOR FILING DATE: 2001-10-26
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 2320
; TYPE: PRT
; ORGANISM: Homo sapiens
```

US-10-279-733-8

```
Query Match      61.5%; Score 1474; DB 14; Length 2320;
Best Local Similarity 46.8%; Pred. No. 1.6e-109;
Matches 351; Conservative 28; Mismatches 75; Indels 296; Gaps 17;

QY      1 PTDLRFTNIGPDTMRVTWAPPSIDLTLNPLVRYSPVKNEEDVAELSISSDNVAVLTNLL 60
DB      1235 PTDLRFTNIGPDTMRVTWAPPSIDLTLNPLVRYSPVKNEEDVAELSISSDNVAVLTNLL 1294

QY      61 PGTETVSVSSVYEQHESPLGRQKTGLDSTGIDFSDITANSFTVHVIAPRATITGYR 120
DB      1295 PGTETVSVSSVYEQHESPLGRQKTGLDSTGIDFSDITANSFTVHVIAPRATITGYR 1354

QY      121 IRHHPHFSGRPREDRVPHSRNSITLTLNTPGTETVSVSIVALNGREESPLLIGQOSTVSD 180
DB      1355 IRHHPHFSGRPREDRVPHSRNSITLTLNTPGTETVSVSIVALNGREESPLLIGQOSTVSD 1414

QY      181 VPRDLVVAAPTPTSLISWDAPAVTVRYRITYGETGNSPVQEFVTPGSKSTATISGLK 240
DB      1415 VPRDLVVAAPTPTSLISWDAPAVTVRYRITYGETGNSPVQEFVTPGSKSTATISGLK 1474

QY      241 PGVDYTIIVYAVTGRGDSPASSKPIISINRYTEIDKPS -----M 278
DB      1475 PGVDYTIIVYAVTGRGDSPASSKPIISINRYTEIDKPS -----M 278

QY      279 AAGSITTLTP----- 287
DB      1535 TGYRVTTTPKNGPGPKTKTAGDQTEMTIEGLQFTVEYVSVYAQNPSGESQPLVQTAV 1594

QY      288 -----ALPEDGGSGAPP-PGHFKDPKRL 309
DB      1595 TNIDRPKGLAFTDVDVDSIKIAWESPOGVSRVRYTYSPPEDGIIHELFPAPDGEDTAEL 1654

QY      310 YCKNGG-----FFLRH-----PDGRV 326
DB      1655 QGLRPGSEYTVSVVALHDDMESQPLIGTQSTAIAPTDLKFTQVTPTSLSAQWTPPNVQL 1714

QY      327 DGVR-----EKSDFHIKLOQAEBRGVVSIGVCANRY-----LAKMEDGRLLASK- 372
DB      1715 TGYRVVTPKEKTGPMKEINLAPDSSVVVSGLMVATKYEVSVYALKD---TLTSRPAQG 1771

QY      373 -----CVTDEC-----FFERLESNNVNTYRS-----R 395
DB      1772 VVTTLNVSPRRARVTDATETTTISWRKTETITGTFOVDAPVANGQTPQRTIKPDVR 1831

QY      396 KYT-----SWYVA 403
DB      1832 SYTITGLQPGTDYKIYLTLDNARSSPVVIDASTAIDAPSNLRLFLATTPNSLLVSWQPP 1891

QY      404 LKRTGQY-----KLGSK-----TG--PGOKAILFL-----PM- 428
DB      1892 RARITGYIIKYEKPGSPREVVRPRPGVTEATITGLEPGTEYTIYVIALKNNQKSEPLI 1951

QY      429 -SAASDELQPLVTLPHNPHLHGPEILDVPST 457
DB      1952 GRKKTDELQPLVTLPHNPHLHGPEILDVPST 1981
```

RESULT 8

```
US-10-182-936A-98
; Sequence 98, Application US/10182936A
; Publication NO. US20040038860A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Kristina M.
; APPLICANT: Anisowicz, Anthony
; APPLICANT: Bhat, Bheem
; APPLICANT: Damagnez, Veronique
; APPLICANT: Robinson, John
; APPLICANT: Yavorsky, Paul
; TITLE OF INVENTION: Reagents and Method for Modulating DKK-Mediated Interactions
; FILE REFERENCE: 032796-143
; CURRENT APPLICATION NUMBER: US/10/182,936A
```

```
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: PCT/US02/15982
; PRIOR FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: US 60/291,311
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: US 60/353,058
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: US 60/361,293
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 216
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 98
; LENGTH: 2328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-182-936A-98

Query Match      61.5%; Score 1474; DB 12; Length 2328;
Best Local Similarity 46.8%; Pred. No. 1.6e-109;
Matches 351; Conservative 28; Mismatches 75; Indels 296; Gaps 17;

QY      1 PTDLRFTNIGPDTMRVTWAPPSIDLTLNPLVRYSPVKNEEDVAELSISSDNVAVLTNLL 60
DB      1243 PTDLRFTNIGPDTMRVTWAPPSIDLTLNPLVRYSPVKNEEDVAELSISSDNVAVLTNLL 1302

QY      61 PGTETVSVSSVYEQHESPLGRQKTGLDSTGIDFSDITANSFTVHVIAPRATITGYR 120
DB      1303 PGTETVSVSSVYEQHESPLGRQKTGLDSTGIDFSDITANSFTVHVIAPRATITGYR 1362

QY      121 IRHHPHFSGRPREDRVPHSRNSITLTLNTPGTETVSVSIVALNGREESPLLIGQOSTVSD 180
DB      1363 IRHHPHFSGRPREDRVPHSRNSITLTLNTPGTETVSVSIVALNGREESPLLIGQOSTVSD 1422

QY      181 VPRDLVVAAPTPTSLISWDAPAVTVRYRITYGETGNSPVQEFVTPGSKSTATISGLK 240
DB      1423 VPRDLVVAAPTPTSLISWDAPAVTVRYRITYGETGNSPVQEFVTPGSKSTATISGLK 1482

QY      241 PGVDYTIIVYAVTGRGDSPASSKPIISINRYTEIDKPS -----M 278
DB      1483 PGVDYTIIVYAVTGRGDSPASSKPIISINRYTEIDKPS -----M 278

QY      279 AAGSITTLTP----- 287
DB      1543 TGYRVTTTPKNGPGPKTKTAGDQTEMTIEGLQFTVEYVSVYAQNPSGESQPLVQTAV 1602

QY      288 -----ALPEDGGSGAPP-PGHFKDPKRL 309
DB      1603 TNIDRPKGLAFTDVDVDSIKIAWESPOGVSRVRYTYSPPEDGIIHELFPAPDGEDTAEL 1662

QY      310 YCKNGG-----FFLRH-----PDGRV 326
DB      1663 QGLRPGSEYTVSVVALHDDMESQPLIGTQSTAIAPTDLKFTQVTPTSLSAQWTPPNVQL 1722

QY      327 DGVR-----EKSDFHIKLOQAEBRGVVSIGVCANRY-----LAKMEDGRLLASK- 372
DB      1723 TGYRVVTPKEKTGPMKEINLAPDSSVVVSGLMVATKYEVSVYALKD---TLTSRPAQG 1779

QY      373 -----CVTDEC-----FFERLESNNVNTYRS-----R 395
DB      1780 VVTTLNVSPRRARVTDATETTTISWRKTETITGTFOVDAPVANGQTPQRTIKPDVR 1839

QY      396 KYT-----SWYVA 403
DB      1840 SYTITGLQPGTDYKIYLTLDNARSSPVVIDASTAIDAPSNLRLFLATTPNSLLVSWQPP 1899

QY      404 LKRTGQY-----KLGSK-----TG--PGOKAILFL-----PM- 428
DB      1900 RARITGYIIKYEKPGSPREVVRPRPGVTEATITGLEPGTEYTIYVIALKNNQKSEPLI 1959

QY      429 -SAASDELQPLVTLPHNPHLHGPEILDVPST 457
DB      1960 GRKKTDELQPLVTLPHNPHLHGPEILDVPST 1989
```

```
RESULT 9
US-10-171-311-64
; Sequence 64, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersn, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 64
; LENGTH: 2328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-64

Query Match 61.5%; Score 1474; DB 14; Length 2328;
Best Local Similarity 46.8%; Pred. No. 1.6e-109;
Matches 351; Conservative 28; Mismatches 75; Indels 296; Gaps 17;

QY 1 PTDLRFTNIGDPTMRVTWAPPSSIDLTLNPLVRYSPVKNEEDVAELSISSDNAVLTNLL 60
DB 1243 PTDLRFTNIGDPTMRVTWAPPSSIDLTLNPLVRYSPVKNEEDVAELSISSDNAVLTNLL 1302

QY 61 PGTEYVSVSSVYEQHESTPLRGOKTGLDPSPTGIDFSDITANSFTVHMIAPRATITGYR 120
DB 1303 PGTEYVSVSSVYEQHESTPLRGOKTGLDPSPTGIDFSDITANSFTVHMIAPRATITGYR 1362

QY 121 IRHHPHFSGRPREDRVPHSRNSITLTLNLTGTEYVSVIVALNGREESPLLIGQQSTVSD 180
DB 1363 IRHHPHFSGRPREDRVPHSRNSITLTLNLTGTEYVSVIVALNGREESPLLIGQQSTVSD 1422

QY 181 VPRDLEVAATPTSLLSISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
DB 1423 VPRDLEVAATPTSLLSISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 1482

QY 241 PGVDYITVYAVTGRGDSPASSKPSISINRYTEIDKPS-----M 278
DB 1483 PGVDYITVYAVTGRGDSPASSKPSISINRYTEIDKPSQMOVTDVQDNSISVKWLPSSSPV 1542

QY 279 AAGSITTL----- 287
DB 1543 TGYRVTTTPKNGPGPTKKTAGPDQTEMTEIQLQPTVEYVSVYAQNPGSGESQPLVQTAV 1602

QY 288 ----- 287
DB 1603 TNIDRPKGLAFTDVDSDIKIAMESPQQGVSRVRYTSSPDGSIHELFPAPDGEETAEL 1662

QY 310 YKNGG-----PFLRIH----- 326
DB 1663 QGLRPGSEYTVSVVALHDDMESOPLIGTOSTAIPATDLKFTQVTPSLSAQWTPPNVOL 1722

QY 327 DGYR-----EKSDPHIKLOAEERGVSIVKVCANRY-----LAKEDGRLLASK----- 372
DB 1723 TGYRVVTPKEKTGPMKEINLAPDSSSVVSGMLMVAIKVEVSVALKD---TLTSRPAQ 1779
```

```
QY 373 -----CVTDEC-----FFFERLESNNYNTYS-----R 395
DB 1780 VVTTLENVSPRRARVTDATEITITISWRTKTETITGTFQVDVAPANGQTPQIORTIKPDVR 1839

QY 396 KYT-----SWYVA 403
DB 1840 SYTITGLQPGTDYKIYLYTLNDNARSPVVIDASTAIDAPSNIURFLATTNSLLVSWQPP 1899

QY 404 LKRTGQY-----KLASK-----TG--PGQKAILFL-----PM- 428
DB 1900 RARITGYIIKYKPGSPREVVPVRPPGVTGTEATITGLEGTETIIVIALKNNQKSEPLI 1959

QY 429 -SAASDELQVLVTLPHNLHGPEILDVPST 457
DB 1960 GRKKTDELQVLVTLPHNLHGPEILDVPST 1989

RESULT 10
US-10-236-031B-70
; Sequence 70, Application US/10236031B
; Publication No. US20030219760A1
; GENERAL INFORMATION:
; APPLICANT: Gordon, Gavin J.
; APPLICANT: Jensen, Roderick V.
; APPLICANT: Gullans, Steven R.
; APPLICANT: Bueno, Raphael
; TITLE OF INVENTION: Diagnostic and Prognostic Tests
; FILE REFERENCE: B00801/70265 (JRV/JAV)
; CURRENT APPLICATION NUMBER: US/10/236,031B
; CURRENT FILING DATE: 2002-09-05
; PRIOR APPLICATION NUMBER: US 60/317,389
; PRIOR FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US 60/407,431
; PRIOR FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70
; LENGTH: 2328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-236-031B-70

Query Match 61.5%; Score 1474; DB 15; Length 2328;
Best Local Similarity 46.8%; Pred. No. 1.6e-109;
Matches 351; Conservative 28; Mismatches 75; Indels 296; Gaps 17;

QY 1 PTDLRFTNIGDPTMRVTWAPPSSIDLTLNPLVRYSPVKNEEDVAELSISSDNAVLTNLL 60
DB 1243 PTDLRFTNIGDPTMRVTWAPPSSIDLTLNPLVRYSPVKNEEDVAELSISSDNAVLTNLL 1302

QY 61 PGTEYVSVSSVYEQHESTPLRGOKTGLDPSPTGIDFSDITANSFTVHMIAPRATITGYR 120
DB 1303 PGTEYVSVSSVYEQHESTPLRGOKTGLDPSPTGIDFSDITANSFTVHMIAPRATITGYR 1362

QY 121 IRHHPHFSGRPREDRVPHSRNSITLTLNLTGTEYVSVIVALNGREESPLLIGQQSTVSD 180
DB 1363 IRHHPHFSGRPREDRVPHSRNSITLTLNLTGTEYVSVIVALNGREESPLLIGQQSTVSD 1422

QY 181 VPRDLEVAATPTSLLSISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240
DB 1423 VPRDLEVAATPTSLLSISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 1482

QY 241 PGVDYITVYAVTGRGDSPASSKPSISINRYTEIDKPS-----M 278
DB 1483 PGVDYITVYAVTGRGDSPASSKPSISINRYTEIDKPSQMOVTDVQDNSISVKWLPSSSPV 1542

QY 279 AAGSITTL----- 287
DB 1543 TGYRVTTTPKNGPGPTKKTAGPDQTEMTEIQLQPTVEYVSVYAQNPGSGESQPLVQTAV 1602

QY 288 ----- 287
DB 1603 TNIDRPKGLAFTDVDSDIKIAMESPQQGVSRVRYTSSPDGSIHELFPAPDGEETAEL 1662

QY 310 YKNGG-----PFLRIH----- 326
DB 1663 QGLRPGSEYTVSVVALHDDMESOPLIGTOSTAIPATDLKFTQVTPSLSAQWTPPNVOL 1722

QY 327 DGYR-----EKSDPHIKLOAEERGVSIVKVCANRY-----LAKEDGRLLASK----- 372
DB 1723 TGYRVVTPKEKTGPMKEINLAPDSSSVVSGMLMVAIKVEVSVALKD---TLTSRPAQ 1779
```

Db 1603 TNIDRPGLAFTDVEDSIKIAMESPOGQSRVRYVTYSSPEDGIHELFPADGEDTAEL 1662  
Qy 310 YKNGG-----FFLRH-----PDGRV 326  
Db 1663 QGLRPGSEYTVSVVALHDDMESQPLIGTQSTAIAPAPDLKFTQVTPSLSAQWTPPNVOL 1722  
Qy 327 DGVR-----EKSDPHIKLQQAEBRGVVSIGVCANRY-----LAMKEDGRLLASK----- 372  
Db 1723 TGYRVVTPKKTGPMKEINLAPDSSVVVSLMWATKYEVSVYALKD---TLTSRPAQG 1779  
Qy 373 -----CVTDEC-----FFERLESNNNTYRS-----R 395  
Db 1780 VVTTLENVSPRRARVTDATETITISWRKTETITGQVDVAVPANGQTPQRTIKPDVR 1839  
Qy 396 KYT-----SWVA 403  
Db 1840 SYTITGLQPGTDYKIYLTLDNARSPVVVIDASTAIDAPSNLRFATTPNSLLVSWOPP 1899  
Qy 404 LKRTGQY-----KLGSK-----TG--PGKAILFL-----PM- 428  
Db 1900 RARITGYIIKYEKPGSPREVVRPRPGVTEATITGLEPGTEYTIYVIALKNNOKSEPLI 1959  
Qy 429 -SAASDELQVLTLPHPNLHGPEILDVPST 457  
Db 1960 GRKKTDELQVLTLPHPNLHGPEILDVPST 1989

RESULT 11  
US-10-374-979-98  
; Sequence 98, Application US/10374979  
; Publication No. US20030219793A1  
; GENERAL INFORMATION:  
; APPLICANT: John P. Carulli et al.  
; TITLE OF INVENTION: THE HIGH BONE MASS GENE OF 11q13.3  
; FILE REFERENCE: 032796-021  
; CURRENT APPLICATION NUMBER: US/10/374,979  
; CURRENT FILING DATE: 2003-03-04  
; PRIOR APPLICATION NUMBER: US 09/544,398  
; PRIOR FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 09/543,771  
; PRIOR FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 09/229,319  
; PRIOR FILING DATE: 1999-01-13  
; PRIOR APPLICATION NUMBER: US 60/071,449  
; PRIOR FILING DATE: 1998-01-13  
; PRIOR APPLICATION NUMBER: US 60/105,511  
; PRIOR FILING DATE: 1998-10-23  
; NUMBER OF SEQ ID NOS: 109  
; SEQ ID NO 98  
; LENGTH: 2328  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-374-979-98

Query Match 61.5%; Score 1474; DB 15; Length 2328;  
Best Local Similarity 46.8%; Pred. No. 1.6e-109;  
Matches 351; Conservative 28; Mismatches 75; Indels 296; Gaps 17;

Qy 1 PTDLRFTNIGDPTMRVTWAPPSPIDLTNLFVRYSPVKNEEDVAELSISSPDNAVLTNLL 60  
Db 1243 PTDLRFTNIGDPTMRVTWAPPSPIDLTNLFVRYSPVKNEEDVAELSISSPDNAVLTNLL 1302  
Qy 61 PGTEYVSVSSVVEQHESTPLRGKQKTGLDSPGTGIDFSIDITANSFVHWIAPRATITGYR 120  
Db 1303 PGTEYVSVSSVVEQHESTPLRGKQKTGLDSPGTGIDFSIDITANSFVHWIAPRATITGYR 1362  
Qy 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGCTEYVSVIVALNGREESPLLIGQOSTVSD 180  
Db 1363 IRHHPHFSGRPREDRVPHSRNSITLTNLTGCTEYVSVIVALNGREESPLLIGQOSTVSD 1422  
Qy 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYGTGGNSPVQEFVPGSKSTATISGLK 240  
Db 1423 VPRDLEVAATPTSLISWDAPAVTVRYRITYGTGGNSPVQEFVPGSKSTATISGLK 1482

Qy 241 PGVDYTTTVAVTRGSDPASSKPIISINRYTEIDKPS-----M 278  
Db 1483 PGVDYTTTVAVTRGSDPASSKPIISINRYTEIDKPSQMQVTDVQDINSISVKWLPSSSPV 1542  
Qy 279 AAGSITTL----- 287  
Db 1543 TGYRVVTPKNGRPGTKTAGPDQTEMTIEGLQPTVEYVSVVAQNPSGESQPLVQTAV 1602  
Qy 288 -----ALPEDGGSGAPP--PGHFKDPKRL 309  
Db 1603 TNIDRPGLAFTDVEDSIKIAMESPOGQSRVRYVTYSSPEDGIHELFPADGEDTAEL 1662  
Qy 310 YKNGG-----FFLRH-----PDGRV 326  
Db 1663 QGLRPGSEYTVSVVALHDDMESQPLIGTQSTAIAPAPDLKFTQVTPSLSAQWTPPNVOL 1722  
Qy 327 DGVR-----EKSDPHIKLQQAEBRGVVSIGVCANRY-----LAMKEDGRLLASK----- 372  
Db 1723 TGYRVVTPKKTGPMKEINLAPDSSVVVSLMWATKYEVSVYALKD---TLTSRPAQG 1779  
Qy 373 -----CVTDEC-----FFERLESNNNTYRS-----R 395  
Db 1780 VVTTLENVSPRRARVTDATETITISWRKTETITGQVDVAVPANGQTPQRTIKPDVR 1839  
Qy 396 KYT-----SWVA 403  
Db 1840 SYTITGLQPGTDYKIYLTLDNARSPVVVIDASTAIDAPSNLRFATTPNSLLVSWOPP 1899  
Qy 404 LKRTGQY-----KLGSK-----TG--PGKAILFL-----PM- 428  
Db 1900 RARITGYIIKYEKPGSPREVVRPRPGVTEATITGLEPGTEYTIYVIALKNNOKSEPLI 1959  
Qy 429 -SAASDELQVLTLPHPNLHGPEILDVPST 457  
Db 1960 GRKKTDELQVLTLPHPNLHGPEILDVPST 1989

RESULT 12

US-10-144-194A-104  
; Sequence 104, Application US/10144194A  
; Publication No. US20030215809A1  
; GENERAL INFORMATION:  
; APPLICANT: Origene Technologies Inc  
; TITLE OF INVENTION: Regulated Breast Cancer Genes  
; FILE REFERENCE: 3U 103 R1  
; CURRENT APPLICATION NUMBER: US/10/144,194A  
; CURRENT FILING DATE: 2002-06-12  
; NUMBER OF SEQ ID NOS: 114  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 104  
; LENGTH: 2355  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-144-194A-104

Query Match 61.5%; Score 1474; DB 15; Length 2355;  
Best Local Similarity 46.8%; Pred. No. 1.6e-109;  
Matches 351; Conservative 28; Mismatches 75; Indels 296; Gaps 17;

Qy 1 PTDLRFTNIGDPTMRVTWAPPSPIDLTNLFVRYSPVKNEEDVAELSISSPDNAVLTNLL 60  
Db 1270 PTDLRFTNIGDPTMRVTWAPPSPIDLTNLFVRYSPVKNEEDVAELSISSPDNAVLTNLL 1329  
Qy 61 PGTEYVSVSSVVEQHESTPLRGKQKTGLDSPGTGIDFSIDITANSFVHWIAPRATITGYR 120  
Db 1330 PGTEYVSVSSVVEQHESTPLRGKQKTGLDSPGTGIDFSIDITANSFVHWIAPRATITGYR 1389  
Qy 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTGCTEYVSVIVALNGREESPLLIGQOSTVSD 180  
Db 1390 IRHHPHFSGRPREDRVPHSRNSITLTNLTGCTEYVSVIVALNGREESPLLIGQOSTVSD 1449  
Qy 181 VPRDLEVAATPTSLISWDAPAVTVRYRITYGTGGNSPVQEFVPGSKSTATISGLK 240

```
Db 1450 VPRLEVAATPTSLLSWDAPAVTVYRITYTGETGNSPVQEFVPGSKSTATISGLK 1509
Qy 241 PGVDYTTTAVTGRGDSPASSKPIISINYTEIDKPS-----M 278
Db 1510 PGVDYTTTAVTGRGDSPASSKPIISINYTEIDKPSQMVQVDVQDNSISVKWLPSSSPV 1569
Qy 279 AAGSITLTP-----287
Db 1570 TGRVVTTPKNGPGTKTAGPDQTEMTIEGLQPTVEYVSVVVAQNPGSGESQPLVQTAV 1629
Qy 288 -----ALPEDGSGGAFPP-PGHFKDPKRL 309
Db 1630 TNIDRPKGLAFTDVDVDSIKIAWESPOGQVSRVRYTSSPEDGIHELFPADGGEEDTAEL 1689
Qy 310 YCKNGG-----FFLRIH-----PDGRV 326
Db 1690 QGLRPGSEYTVSVVALHDDMESQPLIGTQSTAIPTAPDLKFTQVTPSLSAQWTPPNVQL 1749
Qy 327 DGVR-----EKSDPHIKLOQAEERGVSIGVCANRY-----LAMEDGRLLASK-----372
Db 1750 TGRVVRVTCKEKTGPMKEINLAPDSSSVVSGMLVATKYEVSVYALKD---TLTSRPAQ 1806
Qy 373 -----CVTDEC-----FFERLESNNYNTYRS-----R 395
Db 1807 VVTTLENVSPRRARVTDATETITISWRTKTETITGQVDVAVPANGQTPQRTIKPDVR 1866
Qy 396 KYT-----KLGSK-----TG--PGOKAILFL-----PM- 428
Db 1867 SYTITGLQPGTDYKIYLYTLNDNARSPVVVIDASTAIDAPSNLRFATTNPSLLVSWQPP 1926
Qy 404 LKRTGQY-----KLGSK-----TG--PGOKAILFL-----PM- 428
Db 1927 RARITGYIIKYKPGSPREVPRPGVTEATITGLEPGTEYTIYVIALKNNQKSEPLI 1986
Qy 429 -SAASDELQVLTLPHPNLHGPEILDVPST 457
Db 1987 GRKKTDELQVLTLPHPNLHGPEILDVPST 2016

RESULT 13
US-10-360-101-235
; Sequence 235, Application US/10360101
; Publication No. US20040009550A1
; GENERAL INFORMATION:
; APPLICANT: Moll, Gert N.
; APPLICANT: Leenhouts, Cornelis J.
; TITLE OF INVENTION: Export and modification of (poly)peptide in the lantibiotic way
; FILE REFERENCE: 2183-5673
; CURRENT APPLICATION NUMBER: US/10/360,101
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: EP 02077060.8
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 309
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 235
; LENGTH: 2355
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: sequence of fibronectin
US-10-360-101-235

Query Match 61.5%; Score 1474; DB 15; Length 2355;
Best Local Similarity 46.8%; Pred. No. 1.6e-109;
Matches 351; Conservative 28; Mismatches 75; Indels 296; Gaps 17;

Qy 1 PTDLRFNIGPDMRVMTWAPPSPSIDLTNFLVRYSPVKNEEDVAELSISSPSDNAVLTNLL 60
Db 1239 PTDLRFNIGPDMRVMTWAPPSPSIDLTNFLVRYSPVKNEEDVAELSISSPSDNAVLTNLL 1298
Qy 61 PGTEYVSVSSVVEQHESTPLRGQKTGLDPSGTIDFSDITANSFTVHVIAPRATITGYR 120
```

```
Db 1299 PGTEYVSVSSVVEQHESTPLRGQKTGLDPSGTIDFSDITANSFTVHVIAPRATITGYR 1358
Qy 121 IRHHPHFSGRPREDRVPHSRNSITLTNLTPTGTEYVSVI VALNGRERSPLLIQOQSTVSD 180
Db 1359 IRHHPHFSGRPREDRVPHSRNSITLTNLTPTGTEYVSVI VALNGRERSPLLIQOQSTVSD 1418
Qy 181 VPRDLEVAATPTSLLSWDAPAVTVYRITYTGETGNSPVQEFVPGSKSTATISGLK 240
Db 1419 VPRDLEVAATPTSLLSWDAPAVTVYRITYTGETGNSPVQEFVPGSKSTATISGLK 1478
Qy 241 PGVDYTTTAVTGRGDSPASSKPIISINYTEIDKPS-----M 278
Db 1479 PGVDYTTTAVTGRGDSPASSKPIISINYTEIDKPSQMVQVDVQDNSISVKWLPSSSPV 1538
Qy 279 AAGSITLTP-----287
Db 1539 TGRVVTTPKNGPGTKTAGPDQTEMTIEGLQPTVEYVSVVVAQNPGSGESQPLVQTAV 1598
Qy 288 -----ALPEDGSGGAFPP-PGHFKDPKRL 309
Db 1599 TNIDRPKGLAFTDVDVDSIKIAWESPOGQVSRVRYTSSPEDGIHELFPADGGEEDTAEL 1658
Qy 310 YCKNGG-----FFLRIH-----PDGRV 326
Db 1659 QGLRPGSEYTVSVVALHDDMESQPLIGTQSTAIPTAPDLKFTQVTPSLSAQWTPPNVQL 1718
Qy 327 DGVR-----EKSDPHIKLOQAEERGVSIGVCANRY-----LAMEDGRLLASK-----372
Db 1719 TGRVVRVTCKEKTGPMKEINLAPDSSSVVSGMLVATKYEVSVYALKD---TLTSRPAQ 1775
Qy 373 -----CVTDEC-----FFERLESNNYNTYRS-----R 395
Db 1776 VVTTLENVSPRRARVTDATETITISWRTKTETITGQVDVAVPANGQTPQRTIKPDVR 1835
Qy 396 KYT-----KLGSK-----TG--PGOKAILFL-----PM- 428
Db 1836 SYTITGLQPGTDYKIYLYTLNDNARSPVVVIDASTAIDAPSNLRFATTNPSLLVSWQPP 1895
Qy 404 LKRTGQY-----KLGSK-----TG--PGOKAILFL-----PM- 428
Db 1896 RARITGYIIKYKPGSPREVPRPGVTEATITGLEPGTEYTIYVIALKNNQKSEPLI 1955
Qy 429 -SAASDELQVLTLPHPNLHGPEILDVPST 457
Db 1956 GRKKTDELQVLTLPHPNLHGPEILDVPST 1985

RESULT 14
US-10-447-161-3
; Sequence 3, Application US/10447161
; Publication No. US20040023314A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Rong-fu
; TITLE OF INVENTION: Mutant Fibronectin and Tumor Metastasis
; FILE REFERENCE: HO-P02484US1
; CURRENT APPLICATION NUMBER: US/10/447,161
; CURRENT FILING DATE: 2003-05-28
; PRIOR APPLICATION NUMBER: 60/383,530
; PRIOR FILING DATE: 2002-05-28
; NUMBER OF SEQ ID NOS: 148
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3
; LENGTH: 2355
; TYPE: PRT
; ORGANISM: Human
US-10-447-161-3

Query Match 61.5%; Score 1474; DB 16; Length 2355;
Best Local Similarity 46.8%; Pred. No. 1.6e-109;
Matches 351; Conservative 28; Mismatches 75; Indels 296; Gaps 17;

Qy 1 PTDLRFNIGPDMRVMTWAPPSPSIDLTNFLVRYSPVKNEEDVAELSISSPSDNAVLTNLL 60
```

```
Db 1270 PTDLRFTNIGPDMRWVTWAPPPSIDLTNLFVRYSPVKNEEDVAELISISPSDNNAVVLTNLL 1329
Qy 61 PCTEYVSVSSVYEQHESPTLRGQKTGLDSDPTGIDFSDITANSFTVHWIAPRATITGYR 120
Db 1330 PCTEYVSVSVYEQHESPTLRGQKTGLDSDPTGIDFSDITANSFTVHWIAPRATITGYR 1389
Qy 121 IRHPEHSGRPREDRVPHSRNSITLTNLTPTGTEYVSVSVALNGREESPLLIQOQSTVSD 180
Db 1390 IRHPEHSGRPREDRVPHSRNSITLTNLTPTGTEYVSVSVALNGREESPLLIQOQSTVSD 1449
Qy 181 VRDLVVAAATPTSLISWDAPAVTVRYRITYGTGNSPVQEVTPVGSKSTATISGLK 240
Db 1450 VRDLVVAAATPTSLISWDAPAVTVRYRITYGTGNSPVQEVTPVGSKSTATISGLK 1509
Qy 241 PGVDYTIIVYATVGRGDSPASSKPIISINRYTEIDKPS-----M 278
Db 1510 PGVDYTIIVYATVGRGDSPASSKPIISINRYTEIDKPSQOMQVTDVQDNSISVKWLPSSSPV 1569
Qy 279 AAGSITTLT----- 287
Db 1570 TGYRVTTTPKNGPGPKTKTAGPDQTEMTIEGLQPTVEYVSVYAQNPSGESQPLVQTAV 1629
Qy 288 -----ALPEDGGSGAPP-PGHFKDPKRL 309
Db 1630 TNIDRPKGLAFTDVDVDSIKIAWESPGQVSRVRYTYGSPEDGIHELFPAPDGEEDTAE 1689
Qy 310 YCKNGG-----FPLRIH-----PDGRV 326
Db 1690 QGLRGSEYTVSVVVALHDDMESQPLIGTQSTAIPTDCLKFTQVTPTSLSAQWTPNVQL 1749
Qy 327 DGVR-----EKSDDPHIKLOQABERGVSIVKVCANRY-----LAWKEDGRLLASK- 372
Db 1750 TGYRVRVTPKEKTGPMKEINLAPDSSVVVSGLMVATKVEVSVYALKD---TLTSRPAQ 1806
Qy 373 -----CWTDEC-----PFERLESNNNTYRS-----R 395
Db 1807 VVTTLNVSPPRRARVTDATETITISWRKTETITGFGQVDVAVPANGOTPIQRTIKPDVR 1866
Qy 429 -SAASDELQVLTLPHPNLHGPEILDVPST 457
Db 1987 GRKKTDELQVLTLPHPNLHGPEILDVPST 2016
```

```
RESULT 15
US-09-961-403-1
; Sequence 1, Application US/09961403
; Publication No. US20030077589A1
; GENERAL INFORMATION:
; APPLICANT: HE-STUMPP, HOLGER
; APPLICANT: HAENDLER, BERNARD
; APPLICANT: KRAETZSCHMAR, JOERN
; APPLICANT: KREFT, BERTHOLT
; APPLICANT: WINTERHAGER, ELKE
; APPLICANT: REGIDOR, PEDRO
; APPLICANT: SCOTTI, SIMONE
; TITLE OF INVENTION: METHOD FOR IN VITRO DIAGNOSIS OF ENDOMETRIOSIS
; FILE REFERENCE: SCH-1789
; CURRENT APPLICATION NUMBER: US/09/961,403
; CURRENT FILING DATE: 2001-09-25
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2386
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
US-09-961-403-1
Query Match 61.5%; Score 1474; DB 10; Length 2386;
Best Local Similarity 46.8%; Pred. No. 1.7e-109;
Matches 351; Conservative 28; Mismatches 75; Indels 296; Gaps 17;
Qy 1 PTDLRFTNIGPDMRWVTWAPPPSIDLTNLFVRYSPVKNEEDVAELISISPSDNNAVVLTNLL 60
Db 1270 PTDLRFTNIGPDMRWVTWAPPPSIDLTNLFVRYSPVKNEEDVAELISISPSDNNAVVLTNLL 1329
Qy 61 PCTEYVSVSVYEQHESPTLRGQKTGLDSDPTGIDFSDITANSFTVHWIAPRATITGYR 120
Db 1330 PCTEYVSVSVYEQHESPTLRGQKTGLDSDPTGIDFSDITANSFTVHWIAPRATITGYR 1389
Qy 121 IRHPEHSGRPREDRVPHSRNSITLTNLTPTGTEYVSVSVALNGREESPLLIQOQSTVSD 180
Db 1390 IRHPEHSGRPREDRVPHSRNSITLTNLTPTGTEYVSVSVALNGREESPLLIQOQSTVSD 1449
Qy 181 VRDLVVAAATPTSLISWDAPAVTVRYRITYGTGNSPVQEVTPVGSKSTATISGLK 240
Db 1450 VRDLVVAAATPTSLISWDAPAVTVRYRITYGTGNSPVQEVTPVGSKSTATISGLK 1509
Qy 241 PGVDYTIIVYATVGRGDSPASSKPIISINRYTEIDKPS-----M 278
Db 1510 PGVDYTIIVYATVGRGDSPASSKPIISINRYTEIDKPSQOMQVTDVQDNSISVKWLPSSSPV 1569
Qy 279 AAGSITTLT----- 287
Db 1570 TGYRVTTTPKNGPGPKTKTAGPDQTEMTIEGLQPTVEYVSVYAQNPSGESQPLVQTAV 1629
Qy 288 -----ALPEDGGSGAPP-PGHFKDPKRL 309
Db 1630 TNIDRPKGLAFTDVDVDSIKIAWESPGQVSRVRYTYGSPEDGIHELFPAPDGEEDTAE 1689
Qy 310 YCKNGG-----FPLRIH-----PDGRV 326
Db 1690 QGLRGSEYTVSVVVALHDDMESQPLIGTQSTAIPTDCLKFTQVTPTSLSAQWTPNVQL 1749
Qy 327 DGVR-----EKSDDPHIKLOQABERGVSIVKVCANRY-----LAWKEDGRLLASK- 372
Db 1750 TGYRVRVTPKEKTGPMKEINLAPDSSVVVSGLMVATKVEVSVYALKD---TLTSRPAQ 1806
Qy 373 -----CWTDEC-----PFERLESNNNTYRS-----R 395
Db 1807 VVTTLNVSPPRRARVTDATETITISWRKTETITGFGQVDVAVPANGOTPIQRTIKPDVR 1866
Qy 396 KYT----- 403
Db 1867 SYTITGLQPGTDYKIYLYTLNDNARSSPVVIDASTAIDAPSNLRFATTPNSLLVSWQPP 1926
Qy 404 LKRTQY-----KLGSK-----TG--PGOKAILFL-----PM- 428
Db 1927 RARITGYIIKYEKPGSPREVVRPRPGVTEATITGLEPGTEYTIYVIALKNNQKSEPLI 1986
Qy 429 -SAASDELQVLTLPHPNLHGPEILDVPST 457
Db 1987 GRKKTDELQVLTLPHPNLHGPEILDVPST 2016
```

Search completed: May 3, 2004, 13:20:49  
Job time : 59.7263 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 13:06:19 ; Search time 19.9005 Seconds  
(without alignments)  
2208.970 Million cell updates/sec

Title: US-09-775-964-5  
Perfect score: 2398  
Sequence: 1 PDLRFTNIGPDMRTWAP.....LVTLPHPNLHGPEILDVPST 457  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1474	61.5	2386	1 FNHU	fibronectin precu
2	1396	58.2	2265	1 FNHU	fibronectin - bovi
3	1345	56.1	2477	2 S14428	fibronectin precu
4	1018.5	42.5	2481	2 A43908	fibronectin - Afri
5	848	35.4	273	2 A28512	fibronectin - chic
6	824.5	34.4	210	2 A32398	basic fibroblast g
7	812	33.9	157	1 GKBOB	basic fibroblast g
8	792.5	33.0	154	2 A31674	basic fibroblast g
9	777.5	32.0	154	2 C37360	basic fibroblast g
10	768	32.0	189	2 S71465	fibronectin - chic
11	764	31.9	146	1 S00185	basic fibroblast g
12	754.5	31.5	164	2 S31622	basic fibroblast g
13	753	31.4	189	2 A48834	basic fibroblast g
14	738	30.8	137	2 I46711	fibroblast growth
15	681	28.4	155	1 A40117	basic fibroblast g
16	613.5	25.6	1020	2 A29355	fibronectin - chic
17	467	19.5	125	2 A32484	basic fibroblast g
18	425.5	17.7	155	1 A60721	acidic fibroblast
19	417.5	17.4	155	2 A60130	acidic fibroblast
20	416.5	17.4	155	1 A33665	acidic fibroblast
21	411.5	17.2	155	2 S04347	acidic fibroblast
22	411.5	17.2	155	2 D37360	acidic fibroblast
23	403.5	16.8	152	2 JH0476	acidic fibroblast
24	402.5	16.8	155	2 JH0055	acidic fibroblast
25	400.5	16.7	155	1 GKBOA	acidic fibroblast
26	378.5	15.8	1356	2 A45445	janusin precursor
27	358	14.9	1353	1 JH0675	restricin precurs
28	338.5	14.1	1746	1 S19694	tenascin precursor
29	334.5	13.9	2201	2 A32160	tenascin-C - human

30 333.5 13.9 929 2 I51027 type XII collagen  
31 329 13.7 1810 1 A23230 tenascin precursor  
32 327 13.6 2019 1 JQ1322 tenascin precursor  
33 309 12.9 3124 2 A40020 collagen alpha 1(X  
34 302.5 12.6 4135 2 T42629 tenascin-X - bovin  
35 297.5 12.4 843 2 A40970 undulin 1 - human  
36 290.5 12.1 1857 2 S31212 collagen alpha 1(X  
37 290.5 12.1 1888 2 S78476 collagen alpha 1(X  
38 289.5 12.1 1747 2 A45974 cytotactin - chick  
39 280.5 11.7 933 2 A31930 tenascin-X precurs  
40 275.5 11.5 3566 1 A40701 tenascin-X - mouse  
41 274 11.4 860 2 I48839 probable tenascin  
42 274 11.4 4006 2 T09070 collagen alpha 1(V  
43 272 11.3 2944 2 A54849 fibroblast growth  
44 270.5 11.3 206 1 TVHUHS fibroblast growth  
45 266 11.1 194 2 I50710 fibroblast growth

ALIGNMENTS

RESULT 1

FNHU  
fibronectin precursor [validated] - human  
N;Alternate names: fibronectin splice form ED-A  
C;Species: Homo sapiens (man)  
C;Date: 27-Nov-1985 #sequence revision 31-Mar-1993 #text change 08-Dec-2000  
C;Accession: A26460; A26284; S03917; A24854; A24476; A91008; A93529; A21011; A90495; A2  
R;Dean, D.C.; Bowlus, C.L.; Bourgeois, S.  
Proc. Natl. Acad. Sci. U.S.A. 84, 1876-1880, 1987  
A;Title: Cloning and analysis of the promoter region of the human fibronectin gene.  
A;Reference number: A26460; MUID:87175578; PMID:3031656  
A;Accession: A26460  
A;Molecule type: DNA  
A;Residues: 1-49 <DEA>  
A;Cross-references: GB:M15801; NID:gl82686; PIDN:AAA53376.1; PID:g553293  
R;Oldberg, A.; Ruoslahti, E.  
J. Biol. Chem. 261, 2113-2116, 1986  
A;Title: Evolution of the fibronectin gene.  
A;Reference number: A26284; MUID:86111901; PMID:3003095  
A;Accession: A26284  
A;Molecule type: DNA  
A;Residues: 1447-1540 <OLD>  
A;Cross-references: GB:M12549; NID:gl82688  
A;Note: the authors translated the codon TTC for residue 1494 as Glu  
R;Paolella, G.; Henchcliffe, C.; Sebastio, G.; Baralle, F.E.  
Nucleic Acids Res. 16, 3545-3557, 1988  
A;Title: Sequence analysis and in vivo expression show that alternative splicing of ED-  
A;Reference number: S00848; MUID:88233940; PMID:3375063  
A;Accession: S03917  
A;Molecule type: DNA  
A;Residues: 1594-1767, 'V', 1769-1783 <PAO>  
A;Cross-references: EMBL:X07718; NID:g31402  
A;Note: the authors translated the codon AAC for residue 1631 as Asp  
R;Vibe-Pedersen, K.; Magnusson, S.; Baralle, F.E.  
FEBS Lett. 207, 287-291, 1986  
A;Title: Donor and acceptor splice signals within an exon of the human fibronectin gene  
A;Reference number: A24854; MUID:87030929; PMID:3770201  
A;Accession: A24854  
A;Molecule type: DNA  
A;Residues: 1992-2147 <VIB>  
A;Cross-references: GB:X04530; NID:g31436  
R;Gutman, A.; Yamada, K.M.; Kornblitt, A.  
FEBS Lett. 207, 145-148, 1986  
A;Title: Human fibronectin is synthesized as a pre-propolypeptide.  
A;Reference number: A24476; MUID:87030890; PMID:3770189  
A;Accession: A24476  
A;Status: not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 1-14, 'O', 16-38 <GUT>  
R;Kornblitt, A.R.; Umezawa, K.; Vibe-Pedersen, K.; Baralle, F.E.  
EMBO J. 4, 1755-1759, 1985  
A;Title: Primary structure of human fibronectin: differential splicing may generate at

A;Reference number: A91008; MUID:85284965; PMID:2992939  
A;Accession: A91008  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 32-1344;1346-2080;2112-2386 <KOR>  
A;Cross-references: GB:X02761  
R;Kornblitt, A.R.; Vibe-Pedersen, K.; Baralle, F.E.  
Nucleic Acids Res. 12, 5853-5868, 1984  
A;Title: Human fibronectin: cell specific alternative mRNA splicing generates polypeptide  
A;Reference number: A93529; MUID:84272258; PMID:6462919  
A;Accession: A93529  
A;Molecule type: mRNA  
A;Residues: 973-2080;2112-2386 <KOR2>  
A;Cross-references: GB:X00739  
R;Oldberg, A.; Linney, E.; Ruoslahti, E.  
J. Biol. Chem. 258, 10193-10196, 1983  
A;Title: Molecular cloning and nucleotide sequence of a cDNA clone coding for the cell a  
A;Reference number: A21011; MUID:83290929; PMID:6688418  
A;Accession: A21011  
A;Molecule type: mRNA  
A;Residues: 1434-1537 <OL2>  
A;Cross-references: GB:K00055; NID:g182680; PIDN:AAA52459.1; PID:g182683  
R;Bernard, M.F.; Kolbe, M.; Weil, D.; Chu, M.L.  
Biochemistry 24, 2698-2704, 1985  
A;Title: Human cellular fibronectin: comparison of the carboxyl-terminal portion with ra  
A;Reference number: A90495; MUID:85280409; PMID:2992573  
A;Accession: A90495  
A;Molecule type: mRNA  
A;Residues: 1594-2386 <BER>  
A;Cross-references: GB:M10905; NID:g182696; PIDN:AAA52462.1; PID:g182697  
R;Umezawa, K.; Kornblitt, A.R.; Baralle, F.E.  
FEBS Lett. 186, 31-34, 1985  
A;Title: Isolation and characterization of cDNA clones for human liver fibronectin.  
A;Reference number: A22245; MUID:85231203; PMID:2989004  
A;Accession: A22245  
A;Molecule type: mRNA  
A;Residues: 1975-1991;2017-2039 <UM2>  
A;Cross-references: GB:M27590  
R;Sekiguchi, K.; Klos, A.M.; Kurachi, K.; Yoshitake, S.; Hakomori, S.  
Biochemistry 25, 4936-4941, 1986  
A;Title: Human liver fibronectin complementary DNAs: identification of two different mes  
A;Reference number: 152394; MUID:87026578; PMID:3021206  
A;Accession: I65273  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1978-1990;2016-2018;2113-2127 <SEK>  
A;Cross-references: GB:M14060; NID:g182701; PIDN:AAA52464.1; PID:g182704  
R;Kornblitt, A.R.; Vibe-Pedersen, K.; Baralle, F.E.  
Proc. Natl. Acad. Sci. U.S.A. 80, 3218-3222, 1983  
A;Title: Isolation and characterization of cDNA clones for human and bovine fibronectins  
A;Reference number: A21165; MUID:83221567; PMID:6304699  
A;Accession: A21165  
A;Molecule type: mRNA  
A;Residues: 2291-2386 <KOR3>  
A;Cross-references: GB:K00799; NID:g182681; PIDN:AAA52460.1; PID:g182684  
R;Garcia-Pardo, A.; Pearlstein, E.; Frangione, B.  
J. Biol. Chem. 258, 12670-12674, 1983  
A;Title: Primary structure of human plasma fibronectin.  
A;Reference number: A92398; MUID:84032463; PMID:6630202  
A;Accession: A92398  
A;Molecule type: protein  
A;Residues: 32-47, 'C', 49-51, 'S', 53-72, 'A', 74-290 <GAR1>  
R;Garcia-Pardo, A.; Gold, L.I.  
Arch. Biochem. Biophys. 304, 181-188, 1993  
A;Title: Further characterization of the binding of fibronectin to gelatin reveals the F  
A;Reference number: S34791; MUID:93312001; PMID:8323285  
A;Accession: S34791  
A;Molecule type: protein  
A;Residues: 291-300;551-560 <GAR2>

R;Griffin, C.A.; Calaycay, J.; Shively, J.E.; Smith, R.L.  
Thromb. Res. 43, 469-477, 1986  
A;Title: Two plasma fibronectin fragments with different gelatin-binding properties.  
A;Reference number: A60904; MUID:87019725; PMID:3532418  
A;Accession: A60904  
A;Molecule type: protein  
A;Residues: 293-301 <GRI>  
R;Calaycay, J.; Pande, H.; Lee, T.; Borsi, L.; Siri, A.; Shively, J.E.; Zardi, L.  
J. Biol. Chem. 260, 12136-12141, 1985  
A;Title: Primary structure of a DNA- and heparin-binding domain (domain III) in human pl  
A;Reference number: A23901; MUID:86008277; PMID:3900070  
A;Accession: A23901  
A;Molecule type: protein  
A;Residues: 616-677, 'Q', 679-703, 'PT', <CAL>  
R;Pierschbacher, M.D.; Ruoslahti, E.; Sundelin, J.; Lind, P.; Peterson, P.A.  
J. Biol. Chem. 257, 9593-9597, 1982  
A;Title: The cell attachment domain of fibronectin. Determination of the primary structu  
A;Reference number: A92386; MUID:82265604; PMID:7050098  
A;Accession: A92386  
A;Molecule type: protein  
A;Residues: 1441-1548 <PIE>  
A;Note: residues 1524-1527 are responsible for the cell-binding activity  
R;Garcia-Pardo, A.; Rostagno, A.; Frangione, B.  
Biochem. J. 241, 923-928, 1987  
A;Title: Primary structure of human plasma fibronectin. Characterization of a 38 kDa dom  
A;Reference number: A32517; MUID:87241275; PMID:3593230  
A;Accession: A32517  
A;Molecule type: protein  
A;Residues: 1589-1630, 'T', 1722-2058 <GAR3>  
R;Tresselt, T.; McCarthy, J.B.; Calaycay, J.; Lee, T.D.; Legesse, K.; Shively, J.E.; Pand  
Biochem. J. 274, 731-738, 1991  
A;Title: Human plasma fibronectin. Demonstration of structural differences between the A  
A;Reference number: S14357; MUID:91190085; PMID:2012601  
A;Accession: S14357  
A;Molecule type: protein  
A;Residues: 1614-1630, 'T', 1722-2081, 2113-2244 <TRE>  
R;Garcia-Pardo, A.; Pearlstein, E.; Frangione, B.  
J. Biol. Chem. 260, 10320-10325, 1985  
A;Title: Primary structure of human plasma fibronectin. Characterization of a 31,000-dal  
A;Reference number: A23891; MUID:85261459; PMID:4019516  
A;Accession: A23891  
A;Molecule type: protein  
A;Residues: 2071-2080;2112-2356 <GAR4>  
C;Comment: The extra domain and connecting strand 3 are subject to developmental and tis  
C;Comment: The cellular and plasma fibronectins are high molecular weight glycoproteins,  
ation, and transformation.  
C;Genetics:  
A;Gene: GDB:FNI  
A;Cross-references: GDB:119135; OMIM:135600  
A;Map position: 2q34-q34  
A;Introns: 49/3; 1266/1; 1357/1; 1447/1; 1487/1; 1541/1; 1631/1; 1721/1; 1991/1; 2145/1  
C;Superfamily: fibronectin; fibronectin type I repeat homology <IF1>  
F;Keywords: acute phase; alternative splicing; cell adhesion; collagen binding; duplicat  
F;1-26/Domain: signal sequence #status predicted <SIG>  
F;27-31/Domain: propeptide #status predicted <PRO>  
F;32-2386/Product: fibronectin #status experimental <MAT>  
F;52-272/Domain: fibrin and heparin binding <PHB>  
F;52-272/Domain: fibronectin type I repeat homology <IF1>  
F;97-135/Domain: fibronectin type I repeat homology <IF2>  
F;141-179/Domain: fibronectin type I repeat homology <IF3>  
F;186-225/Domain: fibronectin type I repeat homology <IF4>  
F;231-270/Domain: fibronectin type I repeat homology <IF5>  
F;308-608/Domain: collagen binding <CB>  
F;308-342/Domain: fibronectin type I repeat homology <IF6>  
F;360-401/Domain: fibronectin type II repeat homology <2FI>  
F;420-461/Domain: fibronectin type II repeat homology <2F2>  
F;470-508/Domain: fibronectin type I repeat homology <IF7>  
F;518-555/Domain: fibronectin type I repeat homology <IF8>  
F;561-599/Domain: fibronectin type I repeat homology <IF9>  
F;609-692/Domain: fibronectin type III repeat homology <3FA>  
F;616-706/Domain: heparin binding <HPB>  
F;719-801/Domain: fibronectin type III repeat homology <3FB>  
F;810-891/Domain: fibronectin type III repeat homology <3FC>

F;906-988/Domain: fibronectin type III repeat homology <3PD>  
F;996-1077/Domain: fibronectin type III repeat homology <3FE>  
F;1086-1164/Domain: fibronectin type III repeat homology <3FF>  
F;1173-1258/Domain: fibronectin type III repeat homology <3FG>  
F;1266-1349/Domain: fibronectin type III repeat homology <3FH>

Query Match 61.5%; Score 1474; DB 1; Length 2386;  
Best Local Similarity 46.8%; Pred. No. 3.7e-97;  
Matches 351; Conservative 28; Mismatches 75; Indels 296; Gaps 17;

Qy 1 PTDLRFNIGPDTNRVWAPPPSIDLTLNFLVRYSFVKNEEDVAELSLSPSDNAVLTNLL 60  
Db 1270 PTDLRFNIGPDTNRVWAPPPSIDLTLNFLVRYSFVKNEEDVAELSLSPSDNAVLTNLL 1329

Qy 61 PGTEYVVSVSVEQHESTELRGKQTKGLDSPTGIDFSDITANSFTVHVIAPRATITGYR 120  
Db 1330 PGTEYVVSVSVEQHESTELRGKQTKGLDSPTGIDFSDITANSFTVHVIAPRATITGYR 1389

Qy 121 IRHHPHFSGRPREDRVPHSRNSITLNLTPGTEYVVSIVVALNGREESPLLIGQOSTVSD 180  
Db 1390 IRHHPHFSGRPREDRVPHSRNSITLNLTPGTEYVVSIVVALNGREESPLLIGQOSTVSD 1449

Qy 181 VPRDLVVAAATPTSLLSISWAPVTVVYRITYTGETGNSPVQEFVPGSKSTATISGLK 240  
Db 1450 VPRDLVVAAATPTSLLSISWAPVTVVYRITYTGETGNSPVQEFVPGSKSTATISGLK 1509

Qy 241 PGVDYTTIVVAVTGRGDSPASSKPIISINRYTEIDKPS-----M 278  
Db 1510 PGVDYTTIVVAVTGRGDSPASSKPIISINRYTEIDKPSQMQVTDVQDINSISVKWLPSSPV 1569

Qy 279 AAGSITLPL----- 287  
Db 1570 TGRVTVTTTPKNGPGTKTAGPDQTEMTIEGLQPTVEYVVSVAQNPGSESQPLVQTAV 1629

Qy 288 -----ALPDGSGGAPP--PGHFKQPKRL 309  
Db 1630 TNIDRPKGLAFTDVDSISKIAMESPGQVSRVRYTVSSPDGHELFPAAPGDEEDFAEL 1689

Qy 310 YCKNGG-----FFLRIR-----PDGRV 326  
Db 1690 QGLRPGSEYTVSVVALHDDMESQLIGTQSTAIPTAPDLKFTQVTPTSLSAQWTPPNVL 1749

Qy 327 DQVR-----EKSDPHIKLOQAEERGVSIGKVCANRY-----LAMKEDGRLLASK---- 372  
Db 1750 TGYRVRVTPKEKTGPMKEINLAPDSSVWVGLMVAPEYVSVYALKD---TLTSRPAQG 1806

Qy 373 -----CVTDEC-----FFERLESNNYTYRS-----R 395  
Db 1807 VVTTLENVSPRRARVTDATETITISWRKTETITGQVDVAPANGQTPQRTIKPDVR 1866

Qy 396 KYT-----GWYVA 403  
Db 1867 SYTITGLQPGTDYKIYLTLDNARSFVVVIDASTAIDAPSNRFLATTNSLLVSNQPP 1926

Qy 404 LKRTGQY-----KLGSK-----TG--PGQKAILFL-----PM- 428  
Db 1927 RARITGVIIKVEKPGSPREVPRPGVTEATITGLEPGTEYTVIYVIALKNQKSEPLI 1986

Qy 429 -SAASDELQVLVTLPHNLHGPEILDVPST 457  
Db 1987 GRXKTDLPQLVTLPHNLHGPEILDVPST 2016

RESULT 2  
FNOB  
fibronectin - bovine  
C;Species: Bos primigenius taurus (cattle)  
C;Date: 31-Dec-1988 #sequence revision 31-Dec-1988 #text\_change 20-Oct-2000  
C;Accession: A26452; B21165; A23292  
R;Skorstengaard, K.; Jensen, M.S.; Sahl, P.; Petersen, T.E.; Magnusson, S.  
Eur. J. Biochem. 161, 441-453, 1986  
A;Title: Complete primary structure of bovine plasma fibronectin.  
A;Reference number: A26452; MUID:87054047; PMID:3780752

A;Accession: A26452  
A;Molecule type: protein  
A;Residues: 1-2265 <SKO>  
R;Kornblitt, A.R.; Vibe-Pedersen, K.; Baralle, F.E.  
Proc. Natl. Acad. Sci. U.S.A. 80, 3218-3222, 1983  
A;Title: Isolation and characterization of cDNA clones for human and bovine fibronectin  
A;Reference number: A21165; MUID:83221567; PMID:6304699  
A;Accession: B21165  
A;Molecule type: mRNA  
A;Residues: 2170-2265 <KOR>  
A;Cross-references: GB:K00800; NID:G163055; PIDN:AAA30521.2; PID:G5713323  
R;Petersen, T.E.; Thogersen, H.C.; Skorstengaard, K.; Vibe-Pedersen, K.; Sahl, P.; Sott  
Proc. Natl. Acad. Sci. U.S.A. 80, 137-141, 1983  
A;Title: Partial primary structure of bovine plasma fibronectin: three types of interna  
A;Reference number: A23292; MUID:83117805; PMID:6218503  
A;Accession: A23292  
A;Molecule type: protein  
A;Residues: 1-16, 'C', 18-20, 'S', 22-432;447-463;1367-1517;1567-1673;2062-2176, 'N', 2178-22  
C;Comment: Cys-1201 and Cys-2015 have free sulphydryl groups.  
C;Comment: The plasma fibronectin molecule consists of two chains, which are connected i  
C;Comment: Fibronectins bind cell surfaces and various compounds including collagen, fi  
aling, and maintenance of cell shape.  
C;Comment: Plasma fibronectin is synthesized by hepatocytes.  
C;Superfamily: fibronectin; fibronectin type I repeat homology; fibronectin type II rep  
C;Keywords: acute phase; alternative splicing; collagen binding; duplication; extracell  
F;21-241/Domain: fibronectin type I repeat homology <FBR>  
F;66-104/Domain: fibronectin type I repeat homology <1F1>  
F;110-148/Domain: fibronectin type I repeat homology <1F2>  
F;155-194/Domain: fibronectin type I repeat homology <1F3>  
F;200-239/Domain: fibronectin type I repeat homology <1F4>  
F;277-577/Domain: collagen binding <CBR>  
F;329-371/Domain: fibronectin type I repeat homology <1F6>  
F;389-430/Domain: fibronectin type II repeat homology <2F1>  
F;439-477/Domain: fibronectin type II repeat homology <2F2>  
F;487-524/Domain: fibronectin type I repeat homology <1F7>  
F;530-568/Domain: fibronectin type I repeat homology <1F8>  
F;578-661/Domain: fibronectin type III repeat homology <1F9>  
F;688-760/Domain: fibronectin type III repeat homology <FN3A>  
F;779-860/Domain: fibronectin type III repeat homology <FN3B>  
F;875-957/Domain: fibronectin type III repeat homology <FN3C>  
F;965-1046/Domain: fibronectin type III repeat homology <FN3E>  
F;1055-1134/Domain: fibronectin type III repeat homology <FN3F>  
F;1142-1227/Domain: fibronectin type III repeat homology <FN3G>  
F;1235-1318/Domain: fibronectin type III repeat homology <FN3H>  
F;1326-1404/Domain: fibronectin type III repeat homology <GN31>  
F;1410-1517/Domain: cell attachment <CAD>  
F;1416-1502/Domain: fibronectin type III repeat homology <FN3J>  
F;1493-1495/Region: cell attachment (R-G-D) motif  
F;1510-1592/Domain: fibronectin type III repeat homology <FN3K>  
F;1600-1870/Domain: heparin binding <HB2>  
F;1600-1682/Domain: fibronectin type III repeat homology <FN3L>  
F;1692-1773/Domain: fibronectin type III repeat homology <FN3M>  
F;1781-1863/Domain: fibronectin type III repeat homology <FN3N>  
F;1970-1972/Region: cell attachment (R-G-D) motif  
F;1982-2062/Domain: fibronectin type III repeat homology <FN3O>  
F;1985-2216/Domain: fibrin binding <FB2>  
F;2085-2124/Domain: fibronectin type I repeat homology <1F10>  
F;2130-2167/Domain: fibronectin type I repeat homology <1F11>  
F;2174-2209/Domain: fibronectin type I repeat homology <1F12>  
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
F;3/Cross-link: isopeptide (Gln) (interchain to Lys N6-amino of fibrin) #status experim  
F;399,497,511,846,976,1213,1987/Binding site: carbohydrate (Asn) (covalent) #status exp  
F;1205,1692/Binding site: carbohydrate (Asn) (covalent) #status absent  
F;1943,1944/Binding site: carbohydrate (Thr) (covalent) #status experimental  
F;2246/Disulfide bonds: interchain (to 2250) #status predicted  
F;2250/Disulfide bonds: interchain (to 2246) #status predicted  
F;2263/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 58.2%; Score 1396; DB 1; Length 2265;

Best Local Similarity 48.6%; Pred. No. 1.4e-91;  
Matches 327; Conservative 38; Mismatches 76; Indels 232; Gaps 17;  
QY 1 PTDLRTNIGPDMRVMTWAPPSSIDITNLFVYSPVKNEEDVAELISPSDNAVLTNLL 60  
Db 1239 PTDLRTNIGPDMRVMTWAPPSSIELTNLIVRSPVKNEEDVAELISPSDNAVLTNLL 1298  
QY 61 PGTEYVSVSSVVEQHESTPLRGQKTGLDSTGIDFSDITANSTFVHWIAPRAITGYR 120  
Db 1299 PGTEYVSVSVVVEQHESTPLRGQKTALDPSGIDFSDITANSTFVHWIAPRAITGYR 1358  
QY 121 IRRHPHFSGRPREDRVPHSRNSITLNTLPGTEYVSVSIVALNGREESPLLLGQOSTVSD 180  
Db 1359 IRRHPENMGRRPREDRVPHSRNSITLNTLNPGETEYVSVSIVALNSKEESPLVGQOSTVSD 1418  
QY 181 VRDLVVAATPTSLLSISWDAPAVTVYRITYGETGNSPQVQFTVPQSKSTATISGLK 240  
Db 1419 VRDLVIAATPTSLLSISWDAPAVTVYRITYGETGSSPVQVQFTVPQSKSTATISGLK 1478  
QY 241 PGVDYITVYAVTGRGDSPASSKPIGINVTRTEIDKPSMA-----AGSITTLPA---- 288  
Db 1479 PGVDYITVYAVTGRGDSPASSKPIGINVTRTEIDKPSQMVTDVQDNSISVRWLPSSPV 1538  
QY 289 -----LPEDGSGGAFPPGCHFDPK-----RLYCKN----- 313  
Db 1539 TGYRVTTAPKNG-----PGPSKTKVGPDPQTEMTIEGLQPTVEYVSVYAQNQGESQP 1592  
QY 314 -----GGFLRIHPDGRVDGVRKSDPHI 337  
Db 1593 LVQTAVTTPAPTNLKFTQVTPTSLTAQMTAPNVQLTGYRVVTP-----KEKTGPMK 1645  
QY 338 KLOLQAEERGVWSIKGCANRY-----LAMEDGRLASK-----CVT 375  
Db 1646 EINLAPDSSVVVSGLMVATKVSVALKD---TLTSRPAQGVVTTLENVSPPRARVT 1702  
QY 376 DEC-----FFPERLESNNY-----NTYRSRK 396  
Db 1703 DATETTITTSWRTKTTITGFOVDAIPANGQTPQIORTIPDRVSYITITGLQPGTDYKIHL 1762  
QY 397 YT-----SWYVALKKTGOY-----KLGSK 415  
Db 1763 YTLNDNARSPVVIDASTAIDAPSNLRFIATTPNSLLVSNQPPRARITGYIIKYEKPGSP 1822  
QY 416 -----TG--PGQK---AIFLFLPMSAAS-----DELPOLVTLPHP 444  
Db 1823 PREVPRPRPGVTEARITGLEPGETYTIQVIALKNNQKSEPLIGRKKKTDELQVLVTLPHP 1882  
QY 445 NLHGPEILDVPST 457  
Db 1883 NLHGPEILDVPST 1895  
RESULT 3  
S14428  
fibronectin precursor - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 28-Oct-1994 #sequence revision 28-Oct-1994 #text change 20-Aug-1999  
C;Accession: S14428; S14455; A22319; S46203; S00459; A27252; I59049  
R;Hynes, R.O.  
submitted to the EMBL Data Library, July 1989  
A;Reference number: S14428  
A;Accession: S14428  
A;Molecule type: mRNA  
A;Residues: 1-2477 <N>  
A;Cross-references: EMBL:X15906; NID:g56163; PIDN:CAA34020.1; PID:g56164  
R;Schwarzbauer, J.E.; Patel, R.S.; Fonda, D.; Hynes, R.O.  
EMBO J. 6, 2573-2580, 1987  
A;Title: Multiple sites of alternative splicing of the rat fibronectin gene transcript.  
A;Reference number: S12455; MUID:88054951; PMID:2445560  
A;Accession: S12455  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 609-1810,'T',1812-2283 <SCH>

A;Cross-references: EMBL:X15906  
R;Tamkun, J.W.; Schwarzbauer, J.E.; Hynes, R.O.  
Proc. Natl. Acad. Sci. U.S.A. 81, 5140-5144, 1984  
A;Title: A single rat fibronectin gene generates three different mRNAs by alternative sp  
A;Reference number: A22319; MUID:84298097; PMID:6089177  
A;Accession: A22319  
A;Molecule type: DNA  
A;Residues: 2052-2237 <TAM>  
R;Falkenberg, C.; Enghild, J.J.; Thøgersen, I.B.; Salvesen, G.; Akerstrøm, B.  
Biochem. J. 301, 745-751, 1994  
A;Title: Isolation and characterization of fibronectin-alpha(1)-microglobulin complex in  
A;Reference number: S46203; MUID:94330948; PMID:7519849  
A;Accession: S46203  
A;Status: preliminary  
A;Molecule type: protein  
A;Residues: 1183-1192,'GLN',1268,'P',1270-1271,'D',1273,'CF',1276,'PY',1385-1399 <FAL>  
R;Facel, R.S.; Odermatt, E.; Schwarzbauer, J.E.; Hynes, R.O.  
EMBO J. 6, 2565-2572, 1987  
A;Title: Organization of the fibronectin gene provides evidence for exon shuffling durin  
A;Reference number: S00459; MUID:88054950; PMID:3119323  
A;Accession: S00459  
A;Molecule type: DNA  
A;Residues: 1-139;2382-2477 <PAT>  
A;Cross-references: EMBL:X05831  
A;Note: the authors translated the codon CCT for residues 51 and 94 as Ala  
R;Schwarzbauer, J.E.; Tamkun, J.W.; Lemischka, I.R.; Hynes, R.O.  
Cell 35, 421-431, 1983  
A;Title: Three different fibronectin mRNAs arise by alternative splicing within the codi  
A;Reference number: A27252; MUID:84082067; PMID:6317187  
A;Accession: A27252  
A;Molecule type: mRNA  
A;Residues: 1586-1720,'T',1722,1813-2477 <SC2>  
R;Odermatt, E.; Tamkun, J.W.; Hynes, R.O.  
Proc. Natl. Acad. Sci. U.S.A. 82, 6571-6575, 1985  
A;Title: Repeating modular structure of the fibronectin gene: Relationship to protein st  
A;Reference number: I59049; MUID:86016741; PMID:3863113  
A;Accession: I59049  
A;Status: translated from GB/EMBL/DBDJ  
A;Molecule type: DNA  
A;Residues: 1722-1810 <RES>  
A;Cross-references: GB:M11750; NID:g204164; PIDN:AAA41170.1; PID:g554437  
C;Genetics:  
A;Introns: 51/1; 94/1; 2416/3; 2454/3  
C;Superfamily: fibronectin; fibronectin type I repeat homology; fibronectin type II repe  
C;Keywords: alternative splicing; cell adhesion; collagen binding; disulfide bond; duplic  
F;1-3/Domain: signal sequence #status predicted <SIG>  
F;33-2477/Product: fibronectin #status predicted <MAT>  
F;53-88/Domain: fibronectin type I repeat homology <1F1>  
F;98-136/Domain: fibronectin type I repeat homology <1F2>  
F;142-180/Domain: fibronectin type I repeat homology <1F3>  
F;187-226/Domain: fibronectin type I repeat homology <1F4>  
F;232-271/Domain: fibronectin type I repeat homology <1F5>  
F;308-342/Domain: fibronectin type II repeat homology <2F1>  
F;360-401/Domain: fibronectin type II repeat homology <2F2>  
F;420-461/Domain: fibronectin type I repeat homology <1F7>  
F;470-508/Domain: fibronectin type I repeat homology <1F8>  
F;518-555/Domain: fibronectin type I repeat homology <1F9>  
F;561-599/Domain: fibronectin type III repeat homology <FN3A>  
F;609-692/Domain: fibronectin type III repeat homology <FN3B>  
F;718-800/Domain: fibronectin type III repeat homology <FN3C>  
F;809-890/Domain: fibronectin type III repeat homology <FN3D>  
F;905-987/Domain: fibronectin type III repeat homology <FN3E>  
F;995-1076/Domain: fibronectin type III repeat homology <FN3F>  
F;1085-1164/Domain: fibronectin type III repeat homology <FN3G>  
F;1172-1257/Domain: fibronectin type III repeat homology <FN3H>  
F;1265-1348/Domain: fibronectin type III repeat homology <FN3I>  
F;1356-1439/Domain: fibronectin type III repeat homology <FN3J>  
F;1447-1529/Domain: fibronectin type III repeat homology <FN3K>  
F;1537-1619/Domain: fibronectin type III repeat homology <FN3L>  
F;1614-1616/Region: cell attachment (R-G-D) motif  
F;1631-1713/Domain: fibronectin type III repeat homology <FN3M>  
F;1721-1803/Domain: fibronectin type III repeat homology <FN3N>  
F;1811-1893/Domain: fibronectin type III repeat homology <FN3O>

F:1903-1984/Domain: fibronectin type III repeat homology <FN30>  
F:1992-2074/Domain: fibronectin type III repeat homology <FN3P>  
F:2181-2183/Region: cell attachment (R-G-D) motif  
F:2193-2273/Domain: fibronectin type III repeat homology <FN3P>  
F:2296-2335/Domain: fibronectin type I repeat homology <FN3Q>  
F:2341-2378/Domain: fibronectin type I repeat homology <FN3Q>  
F:2385-2420/Domain: fibronectin type I repeat homology <FN3Q>  
F:53-79,77-88,98-126,124-136,142-170,168-180,187-216,232-261,259-271,308-335,333  
368,2368-2378,2385-2411,2409-2420/Disulfide bonds: #status predicted  
F:2458/Disulfide bonds: interchain (to 2462) #status predicted  
F:2462/Disulfide bonds: interchain (to 2458) #status predicted

Query Match 56.1%; Score 1345; DB 2; Length 2477;  
Best Local Similarity 42.4%; Pred. No. 7.1e-98;  
Matches 318; Conservative 49; Mismatches 87; Indels 296; Gaps 18;

Qy 1 PTDLRFTNIGDPTMRVTWAPPSPSIDLTNLFVRYSPVKNEEDVAELSLSPSDNAVLTNLL 60  
Db 1360 PTDLRFTNIGDPTMRVTWAPPSPSIDLTNLFVRYSPVKNEEDVAELSLSPSDNAVLTNLL 1419

Qy 61 PGTEYVSVSVYEQHSTPLRGKQKTGLDSTGIDFSDITANSFTVHVIAPRATITGYR 120  
Db 1420 PGTEYVSVSVYEQHSTPLRGKQKTGLDSTGIDFSDITANSFTVHVIAPRATITGYR 1479

Qy 121 IRHHPHSGRPREDRVPHSRNSITLNTLPGTYVSVVVALNGRESPLLIQOQSTVSD 180  
Db 1480 IRHHPHSGRPREDRVPHSRNSITLNTLPGTYVSVVVALNGRESPLLIQOQSTVSD 1539

Qy 181 VPRDLVVAAPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240  
Db 1540 VPRDLVVAAPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 1599

Qy 241 PGVDYTTVAVTGRGSDPASKKPSISINRYTEIDKPS-----M 278  
Db 1600 PGVDYTTVAVTGRGSDPASKKPSISINRYTEIDKPSQMDVQDMSISVRWLPSTSPV 1559

Qy 279 AAGSITLTP----- 287  
Db 1660 TGYRVTTAPKNGLGFTKSTQVSPDQTEMTIEGLQPTVEYVSVVVAQNRGSESOPVQTAV 1719

Qy 288 -----ALPEDGGSGAPP-PGHPKDPKRL 309  
Db 1720 TNIDRPKGLAFTDVDVDSIKIAMESPOGVSRYRVYSSPEDGHELFAPPDGDEDTAEL 1779

Qy 310 Y-----CKNGGF-----FLRIH-----PDGRV 326  
Db 1780 HGLRPGSEYTVSVVALHGHEMESQPLIGVQSTAIAPNTLNKFTQVSPITLTAQWTAPSKVL 1839

Qy 327 DGVR-----EKSDPHIKLOQAERGVSIGKVCANRY-----LAMKEDGRLASK--- 372  
Db 1840 TGYRVVTPKPKTGPMKEINLSPDSTSVIVSGMLVATKYEVSVVALKD---TLTSRPAQG 1896

Qy 373 -----CVTDEC-----FFERLESNNY----- 389  
Db 1897 VVTTLENVSPRRARVTDATETITISWRTKTEITITGFQVDAIPANGQTPVQRTISPDVR 1956

Qy 390 -----NTYRSRYT-----SRYVA 403  
Db 1957 SYTITGLQPDGYDKHLYTLNDNARSPFVIDASTAIDAPSNLRFLLTTNSLLVSWQAP 2016

Qy 404 LKRTGOY-----KLGSK-----TG-PGQKAILFL-----PM- 428  
Db 2017 RARITGVIIKYKPGSPREVPVRPGVTEATITGLEPGTEYTVIVIALKNNKQKSEPLI 2076

Qy 429 -SAASDELQVLTLPNHLHGPEILLDPST 457  
Db 2077 GRKKTDELQVLTLPNHLHGPEILLDPST 2106

RESULT 4  
A43908  
fibronectin - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)

C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 12-Feb-1999  
C:Accession: A43908  
R:DeSimone, D.W.; Norton, P.A.; Hynes, R.O.  
Dev. Biol. 149, 357-369, 1992  
A:Title: Identification and characterization of alternatively spliced fibronectin mRNAs  
A:Reference number: A43908; MUID:92111942; PMID:1730390  
A:Accession: A43908  
A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-2481 <DES>  
A:Cross-references: GB:M77820  
C:Superfamily: fibronectin; fibronectin type I repeat homology; fibronectin type II rep  
C:Keywords: duplication; extracellular matrix; glycoprotein; heterodimer  
F:100-138/Domain: fibronectin type I repeat homology <FN3>  
F:144-182/Domain: fibronectin type I repeat homology <FN3>  
F:189-228/Domain: fibronectin type I repeat homology <FN3>  
F:234-273/Domain: fibronectin type I repeat homology <FN3>  
F:309-343/Domain: fibronectin type I repeat homology <FN3>  
F:361-402/Domain: fibronectin type II repeat homology <FN3>  
F:421-462/Domain: fibronectin type II repeat homology <FN3>  
F:471-509/Domain: fibronectin type I repeat homology <FN3>  
F:519-556/Domain: fibronectin type I repeat homology <FN3>  
F:562-600/Domain: fibronectin type I repeat homology <FN3>  
F:610-693/Domain: fibronectin type III repeat homology <FN3A>  
F:719-801/Domain: fibronectin type III repeat homology <FN3B>  
F:810-891/Domain: fibronectin type III repeat homology <FN3C>  
F:906-988/Domain: fibronectin type III repeat homology <FN3D>  
F:996-1077/Domain: fibronectin type III repeat homology <FN3E>  
F:1086-1165/Domain: fibronectin type III repeat homology <FN3F>  
F:1173-1258/Domain: fibronectin type III repeat homology <FN3G>  
F:1266-1349/Domain: fibronectin type III repeat homology <FN3H>  
F:1357-1440/Domain: fibronectin type III repeat homology <FN3I>  
F:1448-1530/Domain: fibronectin type III repeat homology <FN3J>  
F:1538-1620/Domain: fibronectin type III repeat homology <FN3K>  
F:1615-1617/Region: cell attachment (R-G-D) motif  
F:1632-1714/Domain: fibronectin type III repeat homology <FN3L>  
F:1722-1804/Domain: fibronectin type III repeat homology <FN3M>  
F:1812-1894/Domain: fibronectin type III repeat homology <FN3N>  
F:1904-1985/Domain: fibronectin type III repeat homology <FN3O>  
F:1993-2075/Domain: fibronectin type III repeat homology <FN3P>  
F:2197-2277/Domain: fibronectin type III repeat homology <FN3Q>  
F:2301-2340/Domain: fibronectin type I repeat homology <FN3R>  
F:2346-2383/Domain: fibronectin type I repeat homology <FN3S>  
F:2390-2425/Domain: fibronectin type I repeat homology <FN3T>  
F:55-81,79-90,100-128,126-138,144-172,170-182,189-218,216-228,234-263,261-273,309-336,3  
2373,2371-2383,2390-2416,2414-2425/Disulfide bonds: #status predicted  
F:2459/Disulfide bonds: interchain (to 2463) #status predicted  
F:2463/Disulfide bonds: interchain (to 2459) #status predicted

Query Match 42.5%; Score 1018.5; DB 2; Length 2481;  
Best Local Similarity 68.7%; Pred. No. 1.9e-64;  
Matches 195; Conservative 43; Mismatches 45; Indels 1; Gaps 1;

Qy 1 PTDLRFTNIGDPTMRVTWAPPSPSIDLTNLFVRYSPVKNEEDVAELSLSPSDNAVLTNLL 60  
Db 1361 PTDLRFTNIGDPTMRVTWAPPSPSIDLTNLFVRYSPVKNEEDVAELSLSPSDNAVLTNLL 1420

Qy 61 PGTEYVSVSVYEQHSTPLRGKQKTGLDSTGIDFSDITANSFTVHVIAPRATITGYR 120  
Db 1421 PGTEYVSVSVYEQHSTPLRGKQKTGLDSTGIDFSDITANSFTVHVIAPRATITGYR 1480

Qy 121 IRHHPHSGRPREDRVPHSRNSITLNTLPGTYVSVVVALNGRESPLLIQOQSTVSD 180  
Db 1481 IRHHPHSGRPREDRVPHSRNSITLNTLPGTYVSVVVALNGRESPLLIQOQSTVSD 1540

Qy 181 VPRDLVVAAPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 240  
Db 1541 VPRDLVVAAPTSLISWDAPAVTVRYRITYGETGNSPVQEFVPGSKSTATISGLK 1600

Qy 241 PGVDYTTVAVTGRGSDPASKKPSISINRYTEIDKP-SMAAGSI 283  
Db 241 PGVDYTTVAVTGRGSDPASKKPSISINRYTEIDKP-SMAAGSI 283

Db 1601 PGVSYTITVYAVTGRGDSPASSKPLTIIHKTDVDPIDMAVTDI 1644

RESULT 5

A28512 fibronectin - chicken (fragment)

C:Species: Gallus gallus (chicken)

C>Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 20-Aug-1999

C:Accession: A28512

R:Kubomura, S.; Obara, M.; Karasaki, Y.; Taniguchi, H.; Gotoh, S.; Tsuda, T.; Higashi, K. Biochim. Biophys. Acta 910, 171-181, 1987

A>Title: Genetic analysis of the cell binding domain region of the chicken fibronectin g

A:Reference number: A28512; MUID:88050950; PMID:2823899

A:Accession: A28512

A:Molecule type: DNA

A:Residues: 1-273 <KUB>

A:Cross-references: GB:X0533; NID:963393; PIDN:CAA29781.1; PID:g295716

A>Note: the authors translated the codon CCG for residue 190 as Gin, CAG for residue 243

C:Genetics:

A:Insertions: 90/1; 129/1; 184/1; 236/1

C:Superfamily: fibronectin; fibronectin type I repeat homology; fibronectin type II repe

C:Keywords: alternative splicing; duplication; extracellular matrix; glycoprotein; heter

F:1-82/Domain: fibronectin type III repeat homology (fragment) <FN31>

F:90-172/Domain: fibronectin type III repeat homology <FN3J>

F:167-169/Region: cell attachment (R-G-D) motif

F:184-266/Domain: fibronectin type III repeat homology <FN3K>

Query Match 35.4%; Score 848; DB 2; Length 273;

Best Local Similarity 84.7%; Pred. No. 1.5e-53;

Matches 160; Conservative 15; Mismatches 14; Indels 0; Gaps 0;

QY 89 LDSPTGIDFSDTANSFTVHWIAPRATITGYRIRHHPHFSGRPREDRVPGRNSITLTN 148

Db 1 LDSPTGLDPSDTANSFTVHWIAPRATITGYRIRHHPHGVGRPKEDRVPGRNSITLTN 60

QY 149 LTFGTYYVSVISVALNGREESPLLIQGOOSTVSDVRDLEVVAAPTPTSLISWDAPAVTRY 208

Db 61 LLPGTYYVSVITAVNGREESVPLVGGQTTVSDVRDLEVNPTSPSTLSISWDAPAVTRY 120

QY 209 YRTYGETGNSVPQSFVPGSKSTATISGLKPGVDYTTIVYAVTGRGDSPASSKPIISIN 268

Db 121 YRTYGETGSSVQEFVPGTMSRATITGLKPGVDYTTIVYAVTGRGDSPASSKPVTVT 180

QY 269 YRTEIDKPS 277

Db 181 YKTEIDTPS 189

RESULT 6

A32398 basic fibroblast growth factor precursor, 22.5K form - human

N:Alternate names: bFGF; fibroblast growth factor 2; prostatic growth factor; prostatrop

N:Contains: basic fibroblast growth factor, 18K form

C:Species: Homo sapiens (man)

C>Date: 31-Jul-1989 #sequence\_revision 31-Dec-1993 #text\_change 21-Jul-2000

C:Accession: A32398; A26642; B32878; S00297; A54316; B54316; A33624; A25824; B24

R:Pras, H.; Kaghad, M.; Prats, A.C.; Klagebrun, M.; Lelias, J.M.; Liauzun, P.; Chalon, Proc. Natl. Acad. Sci. U.S.A. 86, 1836-1840, 1989

A>Title: High molecular mass forms of basic fibroblast growth factor are initiated by al

A:Reference number: A32398; MUID:89184522; PMID:2538817

A:Accession: A32398

A:Molecule type: mRNA

A:Residues: 1-210 <PRA>

A:Cross-references: GB:J04513; NID:gl83083; PIDN:AAA52531.1; PID:g459811

R:Shibata, F.; Baird, A.; Florckiewicz, R.Z.

Growth Factors 4, 277-287, 1991

A>Title: Functional characterization of the human basic fibroblast growth factor gene pr

A:Reference number: A61537; MUID:92110035; PMID:1764564

A:Accession: A61537

A:Molecule type: DNA

A:Residues: 1-114 <SHI>

A>Note: authors translated the codon GGA for residue 47 as Ala

R:Kurokawa, T.; Sasada, R.; Iwane, M.; Igarashi, K.

FEBS Lett. 213, 189-194, 1987

A>Title: Cloning and expression of cDNA encoding human basic fibroblast growth factor.

A:Reference number: A26642; MUID:87162468; PMID:2435575

A:Accession: A26642

A:Molecule type: mRNA

A:Residues: 56-210 <KUR>

A:Cross-references: GB:M27968; NID:gl82562; PIDN:AAA52448.1; PID:gl82563

R:Abraham, J.A.; Whang, J.L.; Tumolo, A.; Mergia, A.; Fiddes, J.C.

Cold Spring Harb. Symp. Quant. Biol. 51, 657-668, 1986

A>Title: Human basic fibroblast growth factor: nucleotide sequence, genomic organization

A:Reference number: A30924; MUID:87217066; PMID:3472745

A:Accession: B32878

A:Molecule type: mRNA

A:Residues: 56-210 <ABR>

A>Note: the authors translated the codon GAA for residue 108 as Gly

R:Abraham, J.A.; Whang, J.L.; Tumolo, A.; Mergia, A.; Friedman, J.; Gospodarowicz, D.; F

EMBO J. 5, 2523-2528, 1986

A>Title: Human basic fibroblast growth factor: nucleotide sequence and genomic organizat

A:Reference number: S00297; MUID:87053817; PMID:3780670

A:Accession: S00297

A>Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-155 <AB2>

A>Note: the authors translated the codon GAA for residue 108 as Gly

R:Shimoyama, Y.; Gotoh, M.; Ino, Y.; Sakamoto, M.; Kato, K.; Hirohashi, S.

Upt. J. Cancer Res. 82, 1263-1270, 1991

A>Title: Characterization of high-molecular-mass forms of basic fibroblast growth factor

rinogenesis.

A:Reference number: A54316; MUID:92091228; PMID:1721615

A:Accession: A54316

A:Molecule type: protein

A:Residues: 'XX', '86-88', 'X', '90-91', 'X', '93-95' <SH3>

A:Experimental source: C-Li21 hepatocellular carcinoma cell line

A>Note: sequence extracted from NCBI backbone (NCBIP:71595)

A:Accession: B54316

A:Molecule type: protein

A:Residues: 'XXX', '19', 'X', '21-29' <SH2>

A:Note: sequence extracted from NCBI backbone (NCBIP:71594)

R:Feige, J.J.; Bradlev, J.D.; Fryburg, K.; Farrie, J.; Cousens, L.C.; Barr, P.J.; Baird, J. Cell Biol. 109, 3105-3114, 1989

A>Title: Differential effects of heparin, fibronectin, and laminin on the phosphorylatic

A:Reference number: A33624; MUID:90078343; PMID:2592418

A:Accession: A33624

A>Status: preliminary

A:Molecule type: protein

A:Residues: 57-210 <FEI>

R:Story, M.T.; Esch, F.; Shimasaki, S.; Sasse, J.; Jacobs, S.C.; Lawson, R.K.

Biochem. Biophys. Res. Commun. 142, 702-709, 1987

A>Title: Amino-terminal sequence of a large form of basic fibroblast growth factor isola

A:Reference number: A25824; MUID:87156686; PMID:2435284

A:Accession: A25824

A:Molecule type: protein

A:Residues: 57-77 <STO>

A:Experimental source: prostate

R:Gimenez-Gallego, G.; Conn, G.; Hatcher, V.B.; Thomas, K.A.

Biochem. Biophys. Res. Commun. 135, 541-548, 1986

A>Title: Human brain-derived acidic and basic fibroblast growth factors: amino terminal

A:Reference number: A30122; MUID:86186784; PMID:3964259

A:Accession: B24243

A:Molecule type: protein

A:Residues: 65-102, 'X', '104-105' <GIM>

A:Experimental source: brain

R:Gautschi, P.; Frater-Schroder, M.; Bohlen, P.

FEBS Lett. 204, 203-207, 1986

A>Title: Partial molecular characterization of endothelial cell mitogens from human brai

A:Reference number: A91364; MUID:86275260; PMID:3732516

A:Accession: B24301

A:Molecule type: protein

A:Residues: 65-88, 'X', '90-98', 'X', '100' <GAU>

R:Sommer, A.; Brewer, M.T.; Thompson, R.C.; Moscatelli, D.; Presta, M.; Rifkin, D.B.

Biochem. Biophys. Res. Commun. 144, 543-550, 1987

A>Title: A form of human basic fibroblast growth factor with an extended amino terminus

A:Reference number: S42242; MUID:87213238; PMID:3579930



A;Accession: S42242  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 54-210 <SOM>  
E;Pantoliano, M.W.; Horlick, R.A.; Springer, B.A.; Van Dyk, D.E.; Tobery, T.; Wetmore, D.  
Biochemistry 33, 10229-10248, 1994  
A;Title: Multivalent ligand-receptor binding interactions in the fibroblast growth factor  
A;Reference number: A55784; PMID:94347757; PMID:7520751  
A;Accession: B55784  
A:Molecule type: protein  
A;Residues: 54-71 <PAN>  
R;Watson, R.; Anthony, F.; Pickett, M.; Lambden, P.; Masson, G.M.; Thomas, E.J.  
Biochem. Biophys. Res. Commun. 187, 1227-1231, 1992  
A;Title: Reverse transcription with nested polymerase chain reaction shows expression of  
clients.

A;Reference number: I52267; PMID:93038590; PMID:1417798  
A;Accession: I52267  
A;Status: preliminary; translated from GB/EMBL/DDBB  
A:Molecule type: mRNA  
A;Residues: 95-182 <RES>  
A;Cross-references: GB:S47380; NID:G956535; PIDN:AAD13853.1; PID:G4261553  
A;Experimental source: granulosa cells  
F;Patry, V.; Bugler, B.; Amalric, F.; Promé, J.C.; Prats, H.  
FEBS Lett. 349, 23-28, 1994

A;Title: Purification and characterization of the 210-amino acid recombinant basic fibro  
A;Reference number: S46253; PMID:94320639; PMID:8045296  
A;Accession: S46253  
A:Molecule type: protein  
A;Residues: 33-53;65-88 <PAT>  
A;Note: recombinant gene expressed in Escherichia coli  
C;Genetics:  
A;Gene: GDB:FGF2; FGFB  
A;Cross-references: GDB:119910; OMIM:134920  
A;Map position: q25-qq27  
A;Start codon: CUG

C;Superfamily: fibroblast growth factor  
C;Keywords: alternative initiators; angiogenesis; growth factor; heparin binding; mitoge  
F;1-210/Product: basic fibroblast growth factor, 22.5K form #status predicted <WA2>  
F;65-210/Product: basic fibroblast growth factor, 18K form #status predicted <MAT>  
F;82-86/Region: heparin binding #status predicted  
F;171-174/Region: heparin binding #status predicted

Query Match 34.4%; Score 824.5; DB 2; Length 210;  
Best Local Similarity 88.8%; Pred. No. 5e-52;  
Matches 159; Conservative 4; Mismatches 15; Indels 1; Gaps 1;

Qy 254 GRGDSPASSKPISINVRTEIDKPSMAAGSTITLPALEDGGSGAFPPGHFKDPRLYCKN 313  
| | : | :  
Db 33 GRGTAAAPRAAPAARGSRPG-PAGTMAAGSTITLPALPEDGGSGAFPPGHFKDPRLYCKN 91  
| | : | :  
Qy 314 GGFFLRHPDGRVDGVREKSDPHIKLOQAERGVSIVKGVCANRYLANKEDEGLLASKC 373  
| | : | :  
Db 92 GGFFLRHPDGRVDGVREKSDPHIKLOQAERGVSIVKGVCANRYLANKEDEGLLASKC 151  
| | : | :  
Qy 374 VTDECFFFFLENNYNTYRSKYTSWYALKRTGYKLGSKTGPQGAILFLPMASAS 432  
| | : | :  
Db 152 VTDECFFFFLENNYNTYRSKYTSWYALKRTGYKLGSKTGPQGAILFLPMASAKS 210  
| | : | :

RESULT 7  
GBKB0B  
basic fibroblast growth factor precursor - bovine (fragment)  
N;Alternate names: bFGF; kidney-derived growth factor; prostattropin  
C;Species: Bos primigenius taurus (cattle)  
C;Date: 13-Aug-1986 #sequence revision 02-Jun-1995 #text change 24-Nov-1999  
C;Accession: A24663; A32878; A33784; A61550; A61551; A60310; A61386; A60316; A22  
R;Abraham, J.A.; Meggia, A.; Whang, J.L.; Tumolo, A.; Friedman, J.; Hjerrild, K.A.; Gosp  
Science 233, 545-548, 1986  
A;Title: Nucleotide sequence of a bovine clone encoding the angiogenic protein, basic fi  
A;Reference number: A94290; PMID:86261806; PMID:2425435  
A;Accession: A24663  
A:Molecule type: mRNA

Proc. Natl. Acad. Sci. U.S.A. 81, 5364-5368, 1984  
A;Title: Isolation and partial molecular characterization of pituitary fibroblast growth  
A;Reference number: A22054; MUID:84298139; PMID:6591194  
A;Accession: A22054  
A;Molecule type: protein  
A;Residues: 12-26 <BOH>  
C;Comment: The acidic and basic fibroblast growth factors are the major endothelial-cell  
cell types in vitro (although bFGF is 30-100 times more potent than aFGF in stimulating t  
C;Comment: This protein binds heparin more strongly than does aFGF.  
C;Superfamily: fibroblast growth factor  
C;Keywords: alternative splicing; angiogenesis; blocked amino end; growth factor; hepari  
F;1-157/Product: basic fibroblast growth factor, uterine form #status predicted <MAT1>  
F;4-157/Product: basic fibroblast growth factor, pituitary gamma form #status experimen  
F;12-157/Product: basic fibroblast growth factor, pituitary alpha form #status experimen  
F;16-157/Product: basic fibroblast growth factor, pituitary short form #status predicted  
F;23-157/Product: basic fibroblast growth factor, hepatic form #status experimental <MAT  
F;29-157/Product: basic fibroblast growth factor, renal form #status experimental <MAT6>  
F;29-33,118-121/Region: heparin binding #status predicted  
F;4/Modified site: blocked amino end (Ala) (in mature form pituitary gamma) (probably ac

Query Match 33.9%; Score 812; DB 1; Length 157;  
Best Local Similarity 97.4%; Pred. No. 2.6e-51;  
Matches 152; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
  
QY 277 SMAAGSITLTPALPEDGGSGAPPGHFKDPKRLYCKNGGFFFLRIHPDGRVDGVREKSDPH 336  
Db 2 AMAAGSITLTPALPEDGGSGAPPGHFKDPKRLYCKNGGFFFLRIHPDGRVDGVREKSDPH 61  
:  
QY 337 IKLQQAERGVSIVKGVCANRYLAMKEDGRLLASKCVTDECFPFERLESNNNTYRSRK 396  
Db 62 IKLQQAERGVSIVKGVCANRYLAMKEDGRLLASKCVTDECFPFERLESNNNTYRSRK 121  
:  
QY 397 YTSWYVALKRTGYKLGSKTGPQKAILFLPMSAAS 432  
Db 122 YSSWYVALKRTGYKLGSKTGPQKAILFLPMSAKS 157  
:  
  
RESULT 8  
A31674  
basic fibroblast growth factor precursor - rat  
N;Alternate names: bFGF  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 21-May-1990 #sequence revision 21-May-1990 #text\_change 16-Jul-1999  
C;Accession: A31674; S00876; S24309  
R;Shimasaki, S.; Emoto, N.; Koba, A.; Mercado, M.; Shibata, F.; Cookesey, K.; Baird, A.;  
Biochem. Biophys. Res. Commun. 157, 256-263, 1988  
A;Title: Complementary DNA cloning and sequencing of rat ovarian basic fibroblast growth  
A;Reference number: A31674; MUID:89061721; PMID:3196337  
A;Accession: A31674  
A;Molecule type: mRNA  
A;Residues: 1-154 <SHI>  
A;Cross-references: GB:M22427; NID:g204285; PIDN:AAA41210.1; PID:g204286  
R;Kurokawa, T.; Seno, M.; Igarashi, K.  
Nucleic Acids Res. 16, 5201, 1988  
A;Title: Nucleotide sequence of rat basic fibroblast growth factor cDNA.  
A;Reference number: S00876; MUID:88462516; PMID:3387229  
A;Accession: S00876  
A;Molecule type: mRNA  
A;Residues: 1-154 <KUR>  
A;Cross-references: EMBL:X07285; NID:g56203; PIDN:CAA30265.1; PID:g56204  
R;El-Husseini, A.E.D.; Paterson, J.A.; Myal, Y.; Shiu, R.P.C.  
Biochim. Biophys. Acta 1131, 314-316, 1992  
A;Title: PCR detection of the rat brain basic fibroblast growth factor (bFGF) mRNA conta  
A;Reference number: S24309; MUID:92329546; PMID:1378302  
A;Accession: S24309  
A;Status: preliminary; translation not shown  
A;Molecule type: mRNA  
A;Residues: 35-154 <ELH>  
A;Cross-references: EMBL:X61697; NID:g56143; PIDN:CAA43863.1; PID:g56144  
C;Superfamily: fibroblast growth factor  
C;Keywords: growth factor  
F;1-9/Domain: signal sequence #status predicted <SIG>  
F;10-154/Product: basic fibroblast growth factor #status predicted <MAT>

Query Match 33.0%; Score 792.5; DB 2; Length 154;  
Best Local Similarity 96.1%; Pred. No. 6.4e-50;  
Matches 149; Conservative 4; Mismatches 1; Indels 1; Gaps 1;  
  
QY 278 MAAGSITLTPALPEDGGSGAPPGHFKDPKRLYCKNGGFFFLRIHPDGRVDGVREKSDPHI 337  
Db 1 MAAGSITSLPALPEDGG-GAPPGHFKDPKRLYCKNGGFFFLRIHPDGRVDGVREKSDPHV 59  
:  
QY 338 KLQQAERGVSIVKGVCANRYLAMKEDGRLLASKCVTDECFPFERLESNNNTYRSRY 397  
Db 60 KLQQAERGVSIVKGVCANRYLAMKEDGRLLASKCVTDECFPFERLESNNNTYRSRY 119  
:  
QY 398 TSWYVALKRTGYKLGSKTGPQKAILFLPMSAAS 432  
Db 120 SSWYVALKRTGYKLGSKTGPQKAILFLPMSAKS 154  
:  
  
RESULT 9  
C37360  
basic fibroblast growth factor - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 17-Apr-1993 #sequence revision 17-Apr-1993 #text\_change 16-Jul-1999  
C;Accession: C37360  
R;Hebert, J.M.; Basilico, C.; Goldfarb, M.; Haub, O.; Martin, G.R.  
Dev. Biol. 138, 454-463, 1990  
A;Title: Isolation of cDNAs encoding four mouse FGF family members and characterization  
A;Reference number: A37360; MUID:90201563; PMID:2318343  
A;Accession: C37360  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-154 <HEB>  
A;Cross-references: GB:M30644; NID:g193296; PIDN:AAA37621.1; PID:g309239  
C;Superfamily: fibroblast growth factor

Query Match 32.4%; Score 777.5; DB 2; Length 154;  
Best Local Similarity 94.2%; Pred. No. 7.7e-49;  
Matches 146; Conservative 5; Mismatches 3; Indels 1; Gaps 1;  
  
QY 278 MAAGSITLTPALPEDGGSGAPPGHFKDPKRLYCKNGGFFFLRIHPDGRVDGVREKSDPHI 337  
Db 1 MAAGSITSLPALPEDGGA-APPGHFKDPKRLYCKNGGFFFLRIHPDGRVDGVREKSDPHV 59  
:  
QY 338 KLQQAERGVSIVKGVCANRYLAMKEDGRLLASKCVTDECFPFERLESNNNTYRSRY 397  
Db 60 KLQQAERGVSIVKGVCANRYLAMKEDGRLLASKCVTDECFPFERLESNNNTYRSRY 119  
:  
QY 398 TSWYVALKRTGYKLGSKTGPQKAILFLPMSAAS 432  
Db 120 SSWYVALKRTGYKLGSKTGPQKAILFLPMSAKS 154  
:  
  
RESULT 10  
S71465  
fibronectin - chicken (fragment)  
C;Species: Gallus gallus (Chicken)  
C;Date: 12-Feb-1998 #sequence revision 13-Mar-1998 #text\_change 21-Jan-2000  
C;Accession: S71465  
R;Gehris, A.L.; Brandli, D.W.; Lewis, S.D.; Bennett, V.D.  
Biochim. Biophys. Acta 1311, 5-12, 1996  
A;Title: The exon encoding the fibronectin type III-9 repeat is constitutively included  
A;Reference number: S71465; MUID:96183658; PMID:8603103  
A;Accession: S71465  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-189 <GEH>  
A;Cross-references: EMBL:U02386; NID:gl323735; PIDN:AAB01062.1; PID:gl323736  
A;Note: the authors translated the codon AAG for residue 50 as Asn and AAG for residue 1  
C;Superfamily: fibronectin; fibronectin type I repeat homology; fibronectin type II repe  
F;100-183/Domain: fibronectin type III repeat homology <3FR>  
  
Query Match 32.0%; Score 768; DB 2; Length 189;  
Best Local Similarity 83.1%; Pred. No. 4.9e-48;





Query Match		30.8%;	Score 738;	DB 2;	Length 137;	
Best Local Similarity		99.3%;	Pred. No. 4.5e-46;			
Matches		136;	Conservative 1;	Mismatches 0;	Indels 0; Gaps 0;	
QY	287	PALPEDGGSGAPP	GHHFKDPKRL	YCKNGGFFLR	IHPDGRVDGVREKSDPHIKLQQAEE	346
Db	1	PALPEDGGSGAPP	GHHFKDPKRL	YCKNGGFFLR	IHPDGRVDGVREKSDPHIKLQQAEE	60
QY	347	GVVSIKGVCANRYL	AMKEDGRLLAS	KCVTDEGFF	FERLESNNYNTYRSRKYTSWYVALKR	406
Db	61	GVVSIKGVCANRYL	AMKEDGRLLAS	KCVTDEGFF	FERLESNNYNTYRSRKYTSWYVALKR	120
QY	407	TGQYKLGSKTGP	QKAI	423		
Db	121	TGQYKLGSKTGP	QKAI	137		
RESULT 15						
A40117						
basic fibroblast growth factor - African clawed frog						
C:Species: Xenopus laevis (African clawed frog)						
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999						
C:Accession: A40117; A29618						
R:Kimelman, D.; Abraham, J.A.; Haaparanta, T.; Palisi, T.M.; Kirschner, M.W.						
Science 242, 1053-1056, 1988						
A:Title: The presence of fibroblast growth factor in the frog egg: its role as a natural						
A:Reference number: A40117; MUID:89058621; PMID:3194757						
A:Accession: A40117						
A>Status: preliminary						
A:Molecule type: mRNA						
A:Residues: 1-155 <KIM>						
A:Cross-references: GB:M18067; NID:G214177; PIDN:AAA49726.1; PID:G214178; GB:M21092						
R:Kimelman, D.; Kirschner, M.						
Cell 51, 869-877, 1987						
A:Title: Synergistic induction of mesoderm by FGF and TGF-beta and the identification of						
A:Reference number: A29618; MUID:88052890; PMID:3479265						
A:Accession: A29618						
A:Molecule type: mRNA						
A:Residues: 95-110,112-155 <KI2>						
C:Superfamily: fibroblast growth factor						
C:Keywords: growth factor						
Query Match		28.4%;	Score 681;	DB 1;	Length 155;	
Best Local Similarity		83.2%;	Pred. No. 6.6e-42;			
Matches		129;	Conservative 9;	Mismatches 17;	Indels 0; Gaps 0;	
QY	278	MAAGSITTLPALPED	GGGAFPPGHH	KDPKRL	YCKNGGFFLR	IHPDGRVDGVREKSDPHI 337
Db	1	MAAGSITTLPE	SGDGNTPFGS	FKDPKRL	YCKNGGFFLR	INSDGRYDGRSDKSDSHI 60
QY	338	KLQQAEE	RGVVISIKGV	CANRYLAMK	EDGRLLAS	KCVTDEGFFFERLESNNYNTYRSKY 397
Db	61	KLQQAEE	RGVVISIKGV	CANRYLAMK	EDGRLLAS	KCVTDEGFFFERLEANNYNTYRSKY 120
QY	398	TSWYVALKRTGQY	KLGSKTGP	QKAILFLP	MSAAS	432
Db	121	SSWYVALKRTGQY	KLGSSTGP	QKAILFLP	MSAKS	155

Result No.	Query Match	Score	Length	DB	ID	Description
1	61.5	1474	2386	1	F1NC HUMAN	P02751 homo sapien
2	58.2	1396	2265	1	F1NC BOVIN	P07589 bos taurus
3	56.1	1345	2477	1	F1NC RAT	P04937 rattus norv
4	55.9	1340	2477	1	F1NC MOUSE	P11276 mus musculus
5	52.10	1210.5	1256	1	F1NC CHICK	P11722 gallus gall
6	43.8	1051.5	1328	1	F1NC PLEWA	Q91289 pleurodeles
7	40.5	1020.5	2481	1	F1NC XENLA	Q91740 xenopus lae
8	42.6	820	155	1	F2G2 HUMAN	P03038 homo sapien
9	34.2	811	155	1	F2G2 BOVIN	P03969 bos taurus
10	33.8	805	155	1	F2G2 SHEEP	P20003 ovis aries
11	33.0	792.5	154	1	F2G2 RAT	P13109 rattus norv
12	32.4	777.5	154	1	F2G2 MOUSE	P15655 mus musculus
13	31.5	754.5	156	1	F2G2 MONDO	P48798 monodelphis
14	31.4	753	158	1	F2G2 CHICK	P48800 gallus gall
15	30.8	738	137	1	F2G2 RABIT	P48799 oryctolagus
16	28.4	681	155	1	F2G2 XENLA	P12226 xenopus lae
17	17.7	425.5	155	1	F2G1 MESAU	P34004 mesocricetu
18	17.4	417.5	155	1	F2G1 CHICK	P19596 gallus gall
19	17.4	416.5	155	1	F2G1 HUMAN	P05230 homo sapien
20	17.2	411.5	155	1	F2G1 MOUSE	P10935 mus musculus
21	16.8	403.5	152	1	F2G1 PIG	P20002 sus scrofa
22	16.7	400.5	155	1	F2G1 BOVIN	P03968 bos taurus
23	14.8	354.5	3063	1	CA1C HUMAN	Q99715 homo sapien
24	14.8	354	1560	1	TENN MOUSE	Q80271 mus musculus
25	14.4	345.5	3119	1	CA1C MOUSE	Q60847 mus musculus
26	14.2	339.5	639	1	CA1C RABIT	Q28902 oryctolagus
27	14.1	338.5	1746	1	TENA PIG	Q29116 sus scrofa
28	13.9	334.5	2201	1	TENA HUMAN	P24821 homo sapien
29	13.9	333.5	929	1	CA1C NOTVI	Q91145 notophthalm
30	13.8	331	1808	1	TENA CHICK	P10039 gallus gall
31	13.6	325	1294	1	TENN HUMAN	Q90cp3 homo sapien
32	12.9	309	522	1	F1NC CANEA	Q28275 canis fami
33	12.9	309	3124	1	CA1C CHICK	P13944 gallus gall

RA Kornblith A.R., Vibe-Pedersen K., Baralle F.E.;  
RT "Human fibronectin: cell specific alternative mRNA splicing generates  
RT polypeptide chains differing in the number of internal repeats.";  
RL Nucleic Acids Res. 12:5853-5868(1984).  
RN [8]  
RP SEQUENCE OF 1232-1782 FROM N.A. (ISOFORM 7).  
RX MEDLINE=88233940; PubMed=3375063;  
RA Paoletti G., Henschcliff C., Sebastio G., Baralle F.E.;  
RT "Sequence analysis and in vivo expression show that alternative  
RT splicing of ED-B and ED-A regions of the human fibronectin gene are  
RT independent events.";  
RL Nucleic Acids Res. 16:3545-3557(1988).  
RN [9]  
RP SEQUENCE OF 1257-1365 FROM N.A. (ISOFORM 11).  
RX MEDLINE=88041070; PubMed=3478690;  
RA Gutman A., Kornblith A.R.;  
RT "Identification of a third region of cell-specific alternative  
RT splicing in human fibronectin mRNA.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:7179-7182(1987).  
RN [10]  
RP SEQUENCE OF 1441-1548.  
RX MEDLINE=82265604; PubMed=7050098;  
RA Pierschbacher M.D., Ruoslahti E., Sundelin J., Lind P., Peterson P.A.;  
RT "The cell attachment domain of fibronectin. Determination of the  
RT primary structure.";  
RL J. Biol. Chem. 257:9593-9597(1982).  
RN [11]  
RP SEQUENCE OF 1448-1540 FROM N.A.  
RX MEDLINE=83290929; PubMed=6688418;  
RA Oldberg A., Linney E., Ruoslahti E.;  
RT "Molecular cloning and nucleotide sequence of a cDNA clone coding for  
RT the cell attachment domain in human fibronectin.";  
RL J. Biol. Chem. 258:10193-10196(1983).  
RN [12]  
RP SEQUENCE OF 1448-1540 FROM N.A.  
RX MEDLINE=86111901; PubMed=3003095;  
RA Oldberg A., Ruoslahti E.;  
RT "Evolution of the fibronectin gene. Exon structure of cell attachment  
RT domain.";  
RL J. Biol. Chem. 261:2113-2116(1986).  
RN [13]  
RP SEQUENCE OF 1594-2386 FROM N.A. (ISOFORM 1).  
RX MEDLINE=85280409; PubMed=2992573;  
RA Bernard M.P., Kolbe M., Weil D., Chu M.-L.;  
RT "Human cellular fibronectin: comparison of the carboxyl-terminal  
RT portion with rat identifies primary structural domains separated by  
RT hypervariable regions.";  
RL Biochemistry 24:2698-2704(1985).  
RN [14]  
RP SEQUENCE OF 1712-1739 FROM N.A.  
RX MEDLINE=87026578; PubMed=3021206;  
RA Sakiguchi K., Kios A.M., Kurachi K., Yoshitake S., Hakomori S.;  
RT "Human liver fibronectin complementary DNAs: identification of two  
RT different messenger RNAs possibly encoding the alpha and beta  
RT subunits of plasma fibronectin.";  
RL Biochemistry 25:4936-4941(1986).  
RN [15]  
RP SEQUENCE OF 1788-2386 FROM N.A. (ISOFORMS 4; 5 AND 6).  
RX TISSUE=Cartilage; PubMed=12127832;  
RA Parker A.E., Boutell J., Carr A., Maciewicz R.A.;  
RT "Novel cartilage-specific splice variants of fibronectin.";  
RL Osteoarthritis Cartilage 10:528-534(2002).  
RN [16]  
RP SEQUENCE OF 32-290.  
RX MEDLINE=84032463; PubMed=6630202;  
RA Garcia-Pardo A., Pearlstein E., Frangione B.;  
RT "Primary structure of human plasma fibronectin. The 29,000-dalton  
RT NH2-terminal domain.";  
RL J. Biol. Chem. 258:12670-12674(1983).  
RN [17]  
RP SEQUENCE OF 309-608, AND COLLAGEN-BINDING.  
RX MEDLINE=87080265; PubMed=3024962;  
RA Owens R.J., Baralle F.E.;  
RT "Mapping the collagen-binding site of human fibronectin by expression  
RT in Escherichia coli.";  
RL EMBO J. 5:2825-2830(1986).  
RN [18]  
RP SULFATION.  
RX MEDLINE=86042625; PubMed=2414772;  
RA Liu M.C., Yu S., Sy J., Redman C.M., Lipmann F.;  
RT "Tyrosine sulfation of proteins from the human hepatoma cell line  
RT HepG2.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:7160-7164(1985).  
RN [19]  
RP O-GLYCOSYLATION OF THR-2064.  
RX MEDLINE=91190085; PubMed=2012601;  
RA Tresselt T., McCarthy J.B., Calaycay J., Lee T.D., Legesse K.,  
RT Shively J.E., Pande H.;  
RT "Human plasma fibronectin. Demonstration of structural differences  
RT between the A- and B-chains in the III CS region.";  
RL Biochem. J. 274:731-738(1991).  
RN [20]  
RP FBLN1-BINDING SITE.  
RX MEDLINE=93015879; PubMed=1400330;  
RA Baibona K., Tran H., Godyna S., Strickland D.K.,  
RT Argraves W.S.;  
RT "Fibulin binds to itself and to the carboxyl-terminal heparin-binding  
RT region of fibronectin.";  
RL J. Biol. Chem. 267:20120-20125(1992).  
RN [21]  
RP CHARACTERIZATION OF FIBRIN-BINDING SITE 1.  
RX MEDLINE=95081153; PubMed=7989369;  
RA Rostagno A., Williams M.J., Baron M., Campbell I.D., Gold L.I.;  
RT "Further characterization of the NH2-terminal fibrin-binding site on  
RT fibronectin.";  
RL J. Biol. Chem. 269:31938-31945(1994).  
RN [22]  
RP STRUCTURE BY NMR OF 1447-1540.  
RX MEDLINE=92162710; PubMed=1311202;  
RA Baron M., Main A.L., Driscoll P.C., Mardon H.J., Boyd J.,  
RT Campbell I.D.;  
RT "1H NMR assignment and secondary structure of the cell adhesion type  
RT III module of fibronectin.";  
RL Biochemistry 31:2068-2073(1992).  
RN [23]  
RP STRUCTURE BY NMR OF 1447-1540.  
RX MEDLINE=93046665; PubMed=1423622;  
RA Main A.L., Harvey T.S., Baron M., Boyd J., Campbell I.D.;  
RT "The three-dimensional structure of the tenth type III module of  
RT fibronectin: an insight into RGD-mediated interactions.";  
RL Cell 71:671-678(1992).  
RN [24]  
RP STRUCTURE BY NMR OF 182-275.  
RX MEDLINE=94141923; PubMed=8308892;  
RA Williams M.J., Phan I., Harvey T.S., Rostagno A., Gold L.I.,  
RA Campbell I.D.;  
RT "Solution structure of a pair of fibronectin type 1 modules with  
RT fibrin binding activity.";  
RL J. Mol. Biol. 235:1302-1311(1994).  
RN [25]  
RP STRUCTURE BY NMR OF 32-92.  
RX MEDLINE=96069779; PubMed=7583666;  
RA Potts J.R., Phan I., Williams M.J., Campbell I.D.;  
RT "High-resolution structural studies of the factor XIIIa crosslinking  
RT site and the first type 1 module of fibronectin.";  
RL Nat. Struct. Biol. 2:946-950(1995).  
RN [26]  
RP STRUCTURE BY NMR OF 406-464.  
RX MEDLINE=98179558; PubMed=9514732;  
RA Sticht H., Pickford A.R., Fotts J.R., Campbell I.D.;  
RT "Solution structure of the glycosylated second type 2 module of  
RT fibronectin.";  
RL J. Mol. Biol. 276:177-187(1998).  
RN [27]  
RP STRUCTURE BY NMR OF EXTRA ED-B DOMAIN FROM ISOFORM 7.

```
Query Match 61.5%, Score 1474, DB 1, Length 2386;
Best Local Similarity 46.8%, Pred. No. 5.4e-96;
Matches 351; Conservative 28; Mismatches 75; Indels 296; Gaps 17;

Qy 1 PTDLRFNIGDTRVTVWAPPSPSIDLNFVLRVSPVKNEDVAELSTSPSDNAVLTNLL 60
Db 1270 PTDLRFNIGDTRVTVWAPPSPSIDLNFVLRVSPVKNEDVAELSTSPSDNAVLTNLL 1329

Qy 61 PGTEYVSVSVYEQHESTPLRGKQKTGLDPSGTIDFSDITANSFTVHWTAPRATITGYR 120
Db 1330 PGTEYVSVSVYEQHESTPLRGKQKTGLDPSGTIDFSDITANSFTVHWTAPRATITGYR 1389

Qy 121 IRHPEHFGSRPREDRVPHSRNSTLTNLTGTEYVSVSVYEQHESTPLRGKQKTGLDPSGTIDFSDITANSFTVHWTAPRATITGYR 180
Db 1390 IRHPEHFGSRPREDRVPHSRNSTLTNLTGTEYVSVSVYEQHESTPLRGKQKTGLDPSGTIDFSDITANSFTVHWTAPRATITGYR 1449

Qy 181 VPRDLEVAATPPTSLISWDAPAVTVRYRITTYGETGNSPVQEFVPGSKSTATISGLK 240
Db 1450 VPRDLEVAATPPTSLISWDAPAVTVRYRITTYGETGNSPVQEFVPGSKSTATISGLK 1509

Qy 241 PGVDYITTVAVTGRGSPASSKPIISINYTEIDKPS-----M 278
Db 1510 PGVDYITTVAVTGRGSPASSKPIISINYTEIDKPSQMVDQDINSISVKWLPSSPV 1569

Qy 279 AAGSITTL----- 287
Db 1570 TGYRVTTTPKNGPPTKTAGDQTEMTIEGLQPTVEYVSVYEQHESTPLRGKQKTGLDPSGTIDFSDITANSFTVHWTAPRATITGYR 1629

Qy 288 -----ALPDGSGGAF-PGHFKDPKRL 309
Db 1630 TNIDRPKGLAFTVDVDSIKIAMESQGVSRVTVYSSPDGHELFPAPEGDEETAEL 1689

Qy 310 YKNGG-----FFLRIH-----PDGRV 326
Db 1690 QGLRPGSEYTVSVVHDDMESQPLIGTQSTAIPTADLKTQVTPSLSAQWTFNVQL 1749

Qy 327 DGVR-----EKSDPHIKLOAEERGVSIGVCANRY-----LAKEDGRLLASK----- 372
Db 1750 TGYRVTVTPKTKGPMKEINLAPDSSSVVGLMVAATKYEVSVVAKD-----TLTSRPAQG 1806

Qy 373 -----CVTDEC-----FFERLESNNYTYRS-----R 395
Db 1807 VTTLENVSPRARVTDATETITISWRTKTETITGQVDAVPANGQTPIQTIKPDVR 1866

Qy 396 KYT-----SWYVA 403
Db 1867 SYTITGLQCTDYKIYLYLNDNARSFVVIDASTAIDAPSNLRFATTPNSLLVSNQPP 1926

Qy 404 LKRTGQY-----KLGSK-----TG--PGQKAILFL-----PM- 428
Db 1927 RARITGYIIKYEKPGSPPREWPRPRPGVTEATITGLEPGCTEYTVIVIALKNNQKSEPLI 1986

Qy 429 -SAASDELQVTLPHNLHGPILDPVST 457
Db 1987 GRKKTDELQVTLPHNLHGPILDPVST 2016
```

## RESULT 2

```
FINC_BOVIN STANDARD; PRT; 2265 AA.
AC P07589;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Fibronectin (FN).
GN FN1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
```

```
RP SEQUENCE.
RX MEDLINE=87054047; PubMed=3780752;
RA Skorstengaard K., Jensen M.S., Sahl P., Petersen T.E., Magnusson S.;
RT "Complete primary structure of bovine plasma fibronectin.";
RN Eur. J. Biochem. 161:441-453(1986).
RP PARTIAL SEQUENCE.
RX MEDLINE=83117805; PubMed=6218503;
RA Petersen T.E., Thorgersen H.C., Skorstengaard K., Vibe-Pedersen K.,
RA Sahl P., Sottrup-Jensen L., Magnusson S.;
RT "Partial primary structure of bovine plasma fibronectin: three types
of internal homology.";
RN Proc. Natl. Acad. Sci. U.S.A. 80:137-141(1983).
RP SEQUENCE OF 2170-2265 FROM N.A.
RX MEDLINE=83221567; PubMed=6304699;
RA Kornblitt A.R., Vibe-Pedersen K., Baralle F.E.;
RT "Isolation and characterization of cDNA clones for human and bovine
fibronectin.";
RN Proc. Natl. Acad. Sci. U.S.A. 80:3218-3222(1983).
CC -!- FUNCTION: Fibronectins bind cell surfaces and various compounds
including collagen, fibrin, heparin, DNA, and actin. Fibronectins
are involved in cell adhesion, cell motility, opsonization, wound
healing, and maintenance of cell shape.
CC -!- SUBUNIT: MOSTLY HETERODIMERS OR MULTIMERS OF ALTERNATIVELY SPLICED
VARIANTS, CONNECTED BY 2 DISULFIDE BONDS NEAR THE CARBOXYL ENDS;
CC TO A LESSER EXTENT HOMODIMERS.
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=1;
CC Comment=A number of isoforms are produced. Each of the "extra
domain" and the connecting strand 3 are present in some forms of
fibronectin and absent in others;
CC Name=1;
CC IsoId=P07589-1; Sequence=Displayed;
CC -!- TISSUE SPECIFICITY: Plasma FN (soluble dimeric form) is secreted
by hepatocytes. Cellular FN (dimeric or cross-linked multimeric
forms), made by fibroblasts, epithelial and other cell types, is
deposited as fibrils in the extracellular matrix.
CC -!- PTM: Sulfated (By similarity).
CC -!- SIMILARITY: Contains 12 fibronectin type I domains.
CC -!- SIMILARITY: Contains 2 fibronectin type II domains.
CC -!- SIMILARITY: Contains 15 fibronectin type III domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: K00800; AAA30521.2; --
CC FIR; A26452; FNBO.
CC HSP; P02751; 2FN2.
CC InterPro: IPR006209; EGF like.
CC InterPro: IPR000083; Fibnctnl.
CC InterPro: IPR008957; FN III-like.
CC InterPro: IPR003961; FN_III.
CC InterPro: IPR000562; FN_Type_II.
CC InterPro: IPR003962; FNIII_subd.
CC Pfam: PF00039; fn1; 12.
CC Pfam: PF00040; fn2; 2.
CC Pfam: PF00041; fn3; 15.
CC PRINTS; PR00013; FNTYPEII.
CC PRINTS; PR00014; FNTYPEIII.
CC ProDom: PD000995; FN_Type_II; 2.
CC SMART; SM00058; FN1; 12.
CC SMART; SM00059; FN2; 2.
CC SMART; SM00060; FN3; 14.
CC PROSITE; PS00022; EGF_1; 2.
CC PROSITE; PS00023; FIBRONECTIN_2; 2.
CC PROSITE; PS01253; FIBRONECTIN_1; 12.
```

KW Glycoprotein; Plasma; Heparin-binding; Acute phase; Phosphorylation;  
KW Sulfation; Cell adhesion; Repeat; Alternative splicing;  
KW Pyridolone carboxylic acid.  
FT MOD\_RES 1 1  
FT DOMAIN 21 241  
FT DOMAIN 277 577  
FT DNAS\_BIND 876 1141  
FT DOMAIN 1236 1509  
FT DOMAIN 1600 1870  
FT DOMAIN 1991 2216  
FT DOMAIN 19 59  
FT DOMAIN 64 107  
FT DOMAIN 108 151  
FT DOMAIN 153 197  
FT DOMAIN 198 242  
FT DOMAIN 275 314  
FT DOMAIN 314 373  
FT DOMAIN 374 438  
FT DOMAIN 437 480  
FT DOMAIN 485 527  
FT DOMAIN 528 571  
FT DOMAIN 578 669  
FT DOMAIN 688 778  
FT DOMAIN 779 874  
FT DOMAIN 875 964  
FT DOMAIN 965 1054  
FT DOMAIN 1055 1141  
FT DOMAIN 1142 1234  
FT DOMAIN 1235 1325  
FT DOMAIN 1326 1415  
FT DOMAIN 1416 1509  
FT DOMAIN 1510 1599  
FT DOMAIN 1600 1691  
FT DOMAIN 1692 1780  
FT DOMAIN 1781 1870  
FT DOMAIN 1871 1990  
FT DOMAIN 1982 2061  
FT DOMAIN 2083 2127  
FT DOMAIN 2128 2170  
FT DOMAIN 2172 2215  
FT SITE 1493 1495  
FT DISULFID 21 47  
FT DISULFID 45 56  
FT DISULFID 66 94  
FT DISULFID 92 104  
FT DISULFID 110 138  
FT DISULFID 136 148  
FT DISULFID 155 184  
FT DISULFID 182 194  
FT DISULFID 200 229  
FT DISULFID 227 239  
FT DISULFID 277 304  
FT DISULFID 302 311  
FT DISULFID 329 355  
FT DISULFID 343 370  
FT DISULFID 389 415  
FT DISULFID 403 430  
FT DISULFID 439 467  
FT DISULFID 465 477  
FT DISULFID 487 514  
FT DISULFID 512 524  
FT DISULFID 530 558  
FT DISULFID 556 568  
FT DISULFID 2085 2114  
FT DISULFID 2112 2124  
FT DISULFID 2130 2157  
FT DISULFID 2155 2167  
FT DISULFID 2174 2200  
FT DISULFID 2198 2209  
FT DISULFID 2246 2246  
FT DISULFID 2250 2250  
FT MOD\_RES 845 845  
FT MOD\_RES 850 850

Glycoprotein; Plasma; Heparin-binding; Acute phase; Phosphorylation;  
Sulfation; Cell adhesion; Repeat; Alternative splicing;  
Pyridolone carboxylic acid.  
MOD\_RES 1 1  
DOMAIN 21 241  
DOMAIN 277 577  
DNAS\_BIND 876 1141  
DOMAIN 1236 1509  
DOMAIN 1600 1870  
DOMAIN 1991 2216  
DOMAIN 19 59  
DOMAIN 64 107  
DOMAIN 108 151  
DOMAIN 153 197  
DOMAIN 198 242  
DOMAIN 275 314  
DOMAIN 314 373  
DOMAIN 374 438  
DOMAIN 437 480  
DOMAIN 485 527  
DOMAIN 528 571  
DOMAIN 578 669  
DOMAIN 688 778  
DOMAIN 779 874  
DOMAIN 875 964  
DOMAIN 965 1054  
DOMAIN 1055 1141  
DOMAIN 1142 1234  
DOMAIN 1235 1325  
DOMAIN 1326 1415  
DOMAIN 1416 1509  
DOMAIN 1510 1599  
DOMAIN 1600 1691  
DOMAIN 1692 1780  
DOMAIN 1781 1870  
DOMAIN 1871 1990  
DOMAIN 1982 2061  
DOMAIN 2083 2127  
DOMAIN 2128 2170  
DOMAIN 2172 2215  
SITE 1493 1495  
DISULFID 21 47  
DISULFID 45 56  
DISULFID 66 94  
DISULFID 92 104  
DISULFID 110 138  
DISULFID 136 148  
DISULFID 155 184  
DISULFID 182 194  
DISULFID 200 229  
DISULFID 227 239  
DISULFID 277 304  
DISULFID 302 311  
DISULFID 329 355  
DISULFID 343 370  
DISULFID 389 415  
DISULFID 403 430  
DISULFID 439 467  
DISULFID 465 477  
DISULFID 487 514  
DISULFID 512 524  
DISULFID 530 558  
DISULFID 556 568  
DISULFID 2085 2114  
DISULFID 2112 2124  
DISULFID 2130 2157  
DISULFID 2155 2167  
DISULFID 2174 2200  
DISULFID 2198 2209  
DISULFID 2246 2246  
DISULFID 2250 2250  
MOD\_RES 845 845  
MOD\_RES 850 850

INTERCHAIN (WITH C-2250).  
INTERCHAIN (WITH C-2246).  
SULFATION (POTENTIAL).  
SULFATION (POTENTIAL).

FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .)  
FT CARBOHYD 497 497 N-LINKED (GLCNAC. . .)  
FT CARBOHYD 511 511 N-LINKED (GLCNAC. . .)  
FT CARBOHYD 846 846 N-LINKED (GLCNAC. . .)  
FT CARBOHYD 976 976 N-LINKED (GLCNAC. . .)  
FT CARBOHYD 1213 1213 N-LINKED (GLCNAC. . .)  
FT CARBOHYD 1987 1987 N-LINKED (GLCNAC. . .)  
FT CARBOHYD 1943 1943 O-LINKED (GALNAC. . .)  
FT CARBOHYD 1944 1944 O-LINKED (GALNAC. . .)  
FT MOD\_RES 2263 2263 PHOSPHORYLATION  
SQ SEQUENCE 2265 AA; 249557 MW; C2D21D486F498D5C CRC64;  
Query Match 58.2%; Score 1396; DB 1; Length 2265;  
Best Local Similarity 48.6%; Pred. No. 1.7e-90;  
Matches 327; Conservative 38; Mismatches 76; Indels 232; Gaps 17;  
QY 1 PTDLRFNIGPDNRVTWAPPPSIDLTFNLYRSPVKNEDVAELSTSPSDNAVLTNLL 60  
DB 1239 PTDLRFNIGPDNRVTWAPPPSIDLTFNLYRSPVKNEDVAELSTSPSDNAVLTNLL 1298  
QY 61 PGTEYVSVSVYBOHESTPLRGQKTGLDPTGIDFSDITANSFTVHWIAPRATITGYR 120  
DB 1299 PGTEYVSVSVYBOHESTPLRGQKTGLDPTGIDFSDITANSFTVHWIAPRATITGYR 1358  
QY 121 IRHPEHFGSRPREDRVPHSRNSITLNTFTGTEYVSVSVYBOHESTPLRGQKTGLDPTGIDFSDITANSFTVHWIAPRATITGYR 180  
DB 1359 IRHPEHFGSRPREDRVPHSRNSITLNTFTGTEYVSVSVYBOHESTPLRGQKTGLDPTGIDFSDITANSFTVHWIAPRATITGYR 1418  
QY 181 VPRDLEVAATPTSLLSWDAVAVTVYRITGTEYVSVSVYBOHESTPLRGQKTGLDPTGIDFSDITANSFTVHWIAPRATITGYR 240  
DB 1419 VPRDLEVAATPTSLLSWDAVAVTVYRITGTEYVSVSVYBOHESTPLRGQKTGLDPTGIDFSDITANSFTVHWIAPRATITGYR 1478  
QY 241 PGVDYTTIVAVTGRGDSPPASSKPIINRYTEIDKPSMA-----AGSTTLTPA----- 288  
DB 1479 PGVDYTTIVAVTGRGDSPPASSKPIINRYTEIDKPSMA-----AGSTTLTPA----- 1538  
QY 289 -----LPDGGSGAPPPGHFKDPK-----RLYCKN----- 313  
DB 1539 TGYRVTAPKNG-----PGSKTKTVGPDQTEMTIEGLQPTVEYVSVYBOHESTPLRGQKTGLDPTGIDFSDITANSFTVHWIAPRATITGYR 1592  
QY 314 -----GGFFLRHDPGRVDGVRKESDPHI 337  
DB 1593 LVQTAVTIPAPTNLKFQVTPTSLTQWAPNVLGTGRVRVP-----KEKTGPMK 1645  
QY 338 KLQQAERGVSIVKGVCANRY-----LAKEDGRLASK-----CVT 375  
DB 1646 EINLAPDSSSVVSLMVATKYEVVYALXD-----TLTSRPAQGVVTTLENVSPRRARVT 1702  
QY 376 DEC-----FFERLESNNY-----NTYRSRK 396  
DB 1703 DATETTTISWRTKTETITGQVDAIPANGQTPQRTIRPDVRSYTTITGLQFGTDYKHL 1762  
QY 397 YT-----SWYVALKRTQY-----KLGSK 415  
DB 1763 YTLNDNARSSPVVIDASTAIDAPSNLRFATPNSLLVSWQPPRARIITYIKYKPGSP 1822  
QY 416 -----TG--PGQK---AIIFLPMSAAS-----DELQVLTPLP 444  
DB 1823 PREVVPRPRPGVTEATITGLRPGTEYTIQVIALKNQKSEPLIGRKTDELQVLTPLP 1882  
QY 445 NLHGPEILDVPST 457  
DB 1883 NLHGPEILDVPST 1895

RESULT 3  
FINC RAT  
ID FINC RAT STANDARD; PRT; 2477 AA.  
AC P04937;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Fibronectin precursor (FN).

GN FN1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Fischer; TISSUE=Liver;  
RX MEDLINE=88054951; PubMed=2445560;  
RA Schwarzbaumer J.E., Patel R.S., Fonda D., Hynes R.O.;  
RT "Multiple sites of alternative splicing of the rat fibronectin gene  
transcript.";  
RL EMBO J. 6:2573-2580(1987).  
RN [2]  
RP SEQUENCE OF 1-139 AND 2382-2477 FROM N.A.  
RC STRAIN=Fischer; TISSUE=Liver;  
RX MEDLINE=88054950; PubMed=3119323;  
RA Patel R.S., Odermatt E., Schwarzbaumer J.E., Hynes R.O.;  
RT "Organization of the fibronectin gene provides evidence for exon  
shuffling during evolution.";  
RL EMBO J. 6:2565-2572(1987).  
RN [3]  
RP SEQUENCE OF 1586-2477 FROM N.A.  
RX MEDLINE=84082067; PubMed=6311787;  
RA Schwarzbaumer J.E., Tamkun J.W., Lemischka I.R., Hynes R.O.;  
RT "Three different fibronectin mRNAs arise by alternative splicing  
within the coding region.";  
RL Cell 35:421-431(1983).  
CC -1- FUNCTION: Fibronectins bind cell surfaces and various compounds  
including collagen, fibrin, heparin, DNA, and actin. Fibronectins  
are involved in cell adhesion, cell motility, opsonization, wound  
healing, and maintenance of cell shape.  
CC -1- SUBUNIT: MOSTLY HETERODIMERS OR MULTIMERS OF ALTERNATIVELY SPLICED  
VARIANTS, CONNECTED BY 2 DISULFIDE BONDS NEAR THE CARBOXYL ENDS;  
CC TO A LESSER EXTENT HOMODIMERS.  
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=4;  
CC Comment=Each of the "extra domain" and the connecting strand 3  
are present in some forms of fibronectin and absent in others;  
CC Name=1;  
CC IsoId=P04937-1; Sequence=Displayed;  
CC Name=2; Synonyms=FNIII-13-16ss;  
CC IsoId=P04937-2; Sequence=VSP\_003258;  
CC Name=3; Synonyms=Lambda-RLF4-5;  
CC IsoId=P04937-3; Sequence=VSP\_003259;  
CC Name=4; Synonyms=Lambda-RLF6;  
CC IsoId=P04937-4; Sequence=VSP\_003260;  
CC -1- TISSUE SPECIFICITY: Plasma FN (soluble dimeric form) is secreted  
by hepatocytes. Cellular FN (dimeric or cross-linked multimeric  
forms), made by fibroblasts, epithelial and other cell types, is  
deposited as fibrils in the extracellular matrix.  
CC -1- PTM: Sulfated (by similarity).  
CC -1- SIMILARITY: Contains 12 fibronectin type I domains.  
CC -1- SIMILARITY: Contains 2 fibronectin type II domains.  
CC -1- SIMILARITY: Contains 17 fibronectin type III domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X15906; CAA34020.1; -;  
DR EMBL; L29191; AAA41166.1; -;  
DR EMBL; L00191; AAA41166.1; JOINED.  
DR EMBL; L29191; AAA41167.1; -;  
DR EMBL; L00191; AAA41167.1; JOINED.  
DR EMBL; L29191; AAA41168.1; -;  
DR EMBL; L00191; AAA41168.1; JOINED.  
DR EMBL; X05831; CAA29278.1; -;

DR EMBL; X05832; CAA29279.1; -;  
DR EMBL; X05833; CAA29280.1; -;  
DR EMBL; X05834; CAA29281.1; -;  
DR PIR; S14428; S14428.  
DR HSP; P02751; IFBR.  
DR InterPro; IPR006209; EGF like.  
DR InterPro; IPR000083; Fibnctnl.  
DR InterPro; IPR008957; FN\_III-like.  
DR InterPro; IPR003961; FN\_III.  
DR InterPro; IPR000562; FN Type II.  
DR InterPro; IPR003962; FNII\_subd.  
DR Pfam; PF00039; fn1; 12.  
DR Pfam; PF00040; fn2; 2.  
DR Pfam; PF00041; fn3; 17.  
DR PRINTS; PR00013; FNTYPEII.  
DR PRINTS; PR00014; FNTYPEIII.  
DR ProDom; PD000995; FN Type II; 2.  
DR SMART; SM00058; FN1\_12.  
DR SMART; SM00059; FN2; 2.  
DR SMART; SM00060; FN3; 13.  
DR PROSITE; PS00022; EGF\_1; 2.  
DR PROSITE; PS00023; FIBRONECTIN 2; 2.  
DR PROSITE; PS01253; FIBRONECTIN 1; 12.  
KW Glycoprotein; Plasma; Heparin-binding; Acute phase; Phosphorylation;  
KW Sulfation; Cell adhesion; Repeat; Alternative splicing; Signal.  
FT SIGNAL 1 32  
FT CHAIN 33 2477 FIBRONECTIN.  
FT DOMAIN 53 273 FIBRIN- AND HEPARIN-BINDING 1.  
FT DOMAIN 308 608 COLLAGEN-BINDING.  
FT DNA\_BIND 906 1171  
FT DOMAIN 1357 1630 CELL-ATTACHMENT.  
FT DOMAIN 1811 2081 HEPARIN-BINDING 2.  
FT DOMAIN 2296 2427 FIBRIN-BINDING 2.  
FT DOMAIN 51 91 FIBRONECTIN TYPE-I 1.  
FT DOMAIN 96 139 FIBRONECTIN TYPE-I 2.  
FT DOMAIN 140 183 FIBRONECTIN TYPE-I 3.  
FT DOMAIN 185 229 FIBRONECTIN TYPE-I 4.  
FT DOMAIN 230 274 FIBRONECTIN TYPE-I 5.  
FT DOMAIN 306 345 FIBRONECTIN TYPE-I 6.  
FT DOMAIN 345 404 FIBRONECTIN TYPE-I 1.  
FT DOMAIN 405 469 FIBRONECTIN TYPE-II 2.  
FT DOMAIN 468 511 FIBRONECTIN TYPE-I 7.  
FT DOMAIN 516 558 FIBRONECTIN TYPE-I 8.  
FT DOMAIN 559 602 FIBRONECTIN TYPE-I 9.  
FT DOMAIN 609 706 FIBRONECTIN TYPE-III 1.  
FT DOMAIN 707 808 FIBRONECTIN TYPE-III 2.  
FT DOMAIN 809 903 FIBRONECTIN TYPE-III 3.  
FT DOMAIN 904 994 FIBRONECTIN TYPE-III 4.  
FT DOMAIN 995 1084 FIBRONECTIN TYPE-III 5.  
FT DOMAIN 1085 1172 FIBRONECTIN TYPE-III 6.  
FT DOMAIN 1173 1264 FIBRONECTIN TYPE-III 7.  
FT DOMAIN 1265 1355 FIBRONECTIN TYPE-III 8 (EXTRA DOMAIN 1).  
FT DOMAIN 1356 1446 FIBRONECTIN TYPE-III 9.  
FT DOMAIN 1447 1536 FIBRONECTIN TYPE-III 10.  
FT DOMAIN 1537 1630 FIBRONECTIN TYPE-III 11.  
FT DOMAIN 1631 1720 FIBRONECTIN TYPE-III 12.  
FT DOMAIN 1721 1810 FIBRONECTIN TYPE-III 13 (EXTRA DOMAIN 2).  
FT DOMAIN 1811 1902 FIBRONECTIN TYPE-III 14.  
FT DOMAIN 1903 1991 FIBRONECTIN TYPE-III 15.  
FT DOMAIN 1992 2081 FIBRONECTIN TYPE-III 16.  
FT DOMAIN 2082 2201 CONNECTING STRAND 3 (CS-3) (V REGION).  
FT DOMAIN 2202 2283 FIBRONECTIN TYPE-III 17.  
FT DOMAIN 2294 2338 FIBRONECTIN TYPE-I 10.  
FT DOMAIN 2339 2381 FIBRONECTIN TYPE-I 11.  
FT DOMAIN 2383 2426 FIBRONECTIN TYPE-I 12.  
FT SITE 1614 1616 CELL ATTACHMENT SITE.  
FT SITE 2181 2183 CELL ATTACHMENT SITE.  
FT DISULFID 53 79 BY SIMILARITY.  
FT DISULFID 77 88 BY SIMILARITY.  
FT DISULFID 98 126 BY SIMILARITY.  
FT DISULFID 124 136 BY SIMILARITY.  
FT DISULFID 142 170 BY SIMILARITY.  
FT DISULFID 168 180 BY SIMILARITY.





RN [3] SEQUENCE OF 562-834 FROM N.A.  
RC STRAIN=NMRI;  
RX MEDLINE=95403556; PubMed=7673336;  
RA Talts J.F., Weller A., Timpl R., Ekblom M., Ekblom P.;  
RT "Regulation of mesenchymal extracellular matrix protein synthesis by  
transforming growth factor-beta and glucocorticoids in tumor  
stroma.";  
RL J. Cell Sci. 108:2153-2162(1995).  
RN [4]  
RN SEQUENCE OF 899-2376 FROM N.A.  
RP Gorski G., Aros M., Norton P.;  
RA Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
RL  
RN [5]  
RN SEQUENCE OF 2375-2477 FROM N.A.  
RP MEDLINE=88124987; PubMed=3124113;  
RA Blatti S.P., Foster D.N., Ranganathan G., Moses H.L., Getz M.J.;  
RT "Induction of fibronectin gene transcription and mRNA is a primary  
response to growth-factor stimulation of AKR-2B cells.";  
RL Proc. Natl. Acad. Sci. U.S.A. 85:1119-1123(1988).  
RN [6]  
RN SEQUENCE OF 2375-2477 FROM N.A.  
RP TISSUE=Kidney;  
RX MEDLINE=93011702; PubMed=1327855;  
RA Khandjian E.W., Salomon C., Leonard N., Tremblay S., Turler H.;  
RT "Fibronectin gene expression in proliferating, quiescent, and SV40-  
infected mouse kidney cells.";  
RL Exp. Cell Res. 202:464-470(1992).  
RN [7]  
RN STRUCTURE BY NMR OF 1447-1630.  
RP MEDLINE=96202578; PubMed=9533887;  
RA Copie V., Tomita Y., Akiyama S.K., Aota S., Yamada K.M., Venable R.M.,  
RA Pastor R.W., Krueger S., Torchia D.A.;  
RT "Solution structure and dynamics of linked cell attachment modules of  
mouse fibronectin containing the RGD and synergy regions: comparison  
with the human fibronectin crystal structure.";  
RL J. Mol. Biol. 277:663-682(1998).  
RN [8]  
RN DOWN-REGULATION BY GLUCOCORTICOIDS.  
RP MEDLINE=21600963; PubMed=11737251;  
RX Gu Y.-C., Talts J.F., Gullberg D., Timpl R., Ekblom M.;  
RA "Glucocorticoids down-regulate the extracellular matrix proteins  
fibronectin, fibulin-1 and fibulin-2 in bone marrow stroma.";  
RL Eur. J. Haematol. 67:176-184(2001).  
RN -|- FUNCTION: Fibronectins bind cell surfaces and various compounds  
including collagen, fibrin, heparin, DNA, and actin. Fibronectins  
are involved in cell adhesion, cell motility, opsonization, wound  
healing, and maintenance of cell shape.  
RN -|- SUBUNIT: Mostly heterodimers or multimers of alternatively spliced  
variants, connected by 2 disulfide bonds near the carboxyl ends;  
to a lesser extend homodimers. Interacts with FBLN1 (By  
similarity).  
RN -|- SUBCELLULAR LOCATION: Secreted; extracellular matrix.  
RN -|- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=1;  
CC Comment=A number of isoforms are produced. Each of the "extra  
domain" and the connecting strand 3 are present in some forms of  
fibronectin and absent in others;  
CC Name=1;  
CC IsoId=P11276-1; Sequence=displayed;  
CC -|- TISSUE SPECIFICITY: Plasma FN (soluble dimeric form) is secreted  
by hepatocytes. Cellular FN (dimeric or cross-linked multimeric  
forms), made by fibroblasts, epithelial and other cell types, is  
deposited as fibrils in the extracellular matrix.  
CC -|- INDUCTION: Glucocorticoids suppressed mRNA expression and protein  
synthesis.  
CC -|- PTM: Sulfated (By similarity).  
CC -|- SIMILARITY: Contains 12 fibronectin type I domains.  
CC -|- SIMILARITY: Contains 2 fibronectin type II domains.  
CC -|- SIMILARITY: Contains 17 fibronectin type III domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; BC051082; AAHS1082.1; -;  
DR EMBL; Z22729; CAA80422.1; -;  
DR EMBL; X82402; CAA57796.1; -;  
DR EMBL; X93167; CAA63654.1; -;  
DR EMBL; M18194; AAA37636.1; -;  
DR EMBL; S45680; AAB23491.1; -;  
DR PIR; A49173; A49173;  
DR PIR; I48349; I48349;  
DR PDB; 1MFN; 29-APR-98.  
DR PDB; 2MFN; 29-APR-98.  
DR MGD; MGI:95566; Fnl.  
DR GO; GO:0007155; Pcell adhesion; IDA.  
DR InterPro; IPR006209; EGF like.  
DR InterPro; IPR000083; Fibrctnl.  
DR InterPro; IPR008957; FN III-like.  
DR InterPro; IPR003961; FN III.  
DR InterPro; IPR000562; FN\_Type II.  
DR InterPro; IPR003962; FnIII\_subd.  
DR InterPro; IPR008924; MCR\_alpha\_beta\_C.  
DR PRINTS; PR00014; FNTYPEIII.  
DR SMART; SM00058; FN1; 4.  
DR SMART; SM00060; FN3; 12.  
DR PROSITE; PS00022; EGF\_1; 2.  
DR PROSITE; PS00023; FIBRONECTIN\_2; 2.  
DR PROSITE; PS01253; FIBRONECTIN\_1; 12.  
KW Glycoprotein; Plasma; Heparin-binding; Acute phase; Phosphorylation;  
KW Sulfation; Cell adhesion; Repeat; Alternative splicing; Signal;  
KW 3D-structure.  
FT SIGNAL 1 32 BY SIMILARITY.  
FT CHAIN 33 2477 FIBRONECTIN.  
FT DOMAIN 53 273 FIBRIN- AND HEPARIN-BINDING 1.  
FT DOMAIN 308 608 COLLAGEN-BINDING.  
FT DOMAIN 1357 1630 CELL-ATTACHMENT.  
FT DOMAIN 1811 2081 HEPARIN-BINDING 2.  
FT DOMAIN 2296 2427 FIBRIN-BINDING 2.  
FT DOMAIN 51 96 FIBRONECTIN TYPE-I 1.  
FT DOMAIN 96 140 FIBRONECTIN TYPE-I 2.  
FT DOMAIN 140 185 FIBRONECTIN TYPE-I 3.  
FT DOMAIN 185 230 FIBRONECTIN TYPE-I 4.  
FT DOMAIN 230 272 FIBRONECTIN TYPE-I 5.  
FT DOMAIN 306 343 FIBRONECTIN TYPE-I 6.  
FT DOMAIN 345 404 FIBRONECTIN TYPE-II 1.  
FT DOMAIN 405 469 FIBRONECTIN TYPE-II 2.  
FT DOMAIN 468 516 FIBRONECTIN TYPE-I 7.  
FT DOMAIN 516 559 FIBRONECTIN TYPE-I 8.  
FT DOMAIN 559 602 FIBRONECTIN TYPE-I 9.  
FT DOMAIN 609 706 FIBRONECTIN TYPE-III 1.  
FT DOMAIN 707 808 FIBRONECTIN TYPE-III 2.  
FT DOMAIN 809 903 FIBRONECTIN TYPE-III 3.  
FT DOMAIN 904 994 FIBRONECTIN TYPE-III 4.  
FT DOMAIN 995 1084 FIBRONECTIN TYPE-III 5.  
FT DOMAIN 1085 1172 FIBRONECTIN TYPE-III 6.  
FT DOMAIN 1173 1264 FIBRONECTIN TYPE-III 7.  
FT DOMAIN 1265 1355 FIBRONECTIN TYPE-III 8 (EXTRA DOMAIN 1).  
FT DOMAIN 1356 1446 FIBRONECTIN TYPE-III 9.  
FT DOMAIN 1447 1536 FIBRONECTIN TYPE-III 10.  
FT DOMAIN 1537 1630 FIBRONECTIN TYPE-III 11.  
FT DOMAIN 1631 1720 FIBRONECTIN TYPE-III 12. (EXTRA DOMAIN 2).  
FT DOMAIN 1721 1810 FIBRONECTIN TYPE-III 13.  
FT DOMAIN 1811 1902 FIBRONECTIN TYPE-III 14.  
FT DOMAIN 1903 1991 FIBRONECTIN TYPE-III 15.  
FT DOMAIN 1992 2081 FIBRONECTIN TYPE-III 16.  
FT DOMAIN 2082 2201 CONNECTING STRAND 3 (CS-3) (V REGION).  
FT DOMAIN 2202 2283 FIBRONECTIN TYPE-III 17.  
FT DOMAIN 2284 2338 FIBRONECTIN TYPE-I 10.  
FT DOMAIN 2339 2381 FIBRONECTIN TYPE-I 11.

FT	DOMAIN	2383	2426	FIBRONECTIN TYPE-I 12.
FT	SITE	1614	1616	CELL ATTACHMENT SITE.
FT	SITE	2181	2183	CELL ATTACHMENT SITE.
FT	DISULFID	53	79	BY SIMILARITY.
FT	DISULFID	77	88	BY SIMILARITY.
FT	DISULFID	98	126	BY SIMILARITY.
FT	DISULFID	124	136	BY SIMILARITY.
FT	DISULFID	142	170	BY SIMILARITY.
FT	DISULFID	168	180	BY SIMILARITY.
FT	DISULFID	187	216	BY SIMILARITY.
FT	DISULFID	214	226	BY SIMILARITY.
FT	DISULFID	232	261	BY SIMILARITY.
Query Match 55.9%; Score 1340; DB 1; Length 2477;				
Best Local Similarity 41.7%; Pred. No. 1.7e-86;				
Matches 316; Conservative 46; Mismatches 85; Indels 310; Gaps 16;				
QY	1	PTDLRTNIGPDMRTVMTWAPPSPIDITNLFVSPVKNEDVAELISPSDNAVLTNLL	60	
DB	1360	PTDLRTNIGPDMRTVMTWAPPSPIDITNLFVSPVKNEDVAELISPSDNAVLTNLL	1419	
QY	61	PGTEYVSVSSVYEQHESPLRGOKTGLDPSPTGIDFSDITANSFTVHVIAPRATITGYR	120	
DB	1420	PGTEYLVSVSSVYEQHESPLRGOKTGLDPSPTGIDFSDITANSFTVHVIAPRATITGYI	1479	
QY	121	IRHPHFHSGRPREDVRPHSRNSITITLNTLPGTEYVSVSVIANGREESPLLICQOSTVSD	180	
DB	1480	IRHAEHSVGRPDQDRVPSPRSNITITLNTLPGTEYVSVIANGREESPLLICQOSTVSD	1539	
QY	181	VPDLEWVAATPSLLISHDAPAVTVYRITYRITGETGNSPVOEFTVPGSKSTATISGLK	240	
DB	1540	IPRDLVIASTPTSLISWEPPAVSVYRITYRITGETGNSPVOEFTVPGSKSTATINNIK	1599	
QY	241	PGVDYITTVAVTGRGDSPASSKPI SINVRTETDKPS	278	
DB	1600	PGADYITTVAVTGRGDSPASSKPI SINVRTETDKPSQMVTDVONSISVRWLPSTSPV	1659	
QY	279	AAGSITLTP	287	
DB	1660	TGVRVITTPKNGLGSPKTKTASPDQTEMTIEGLQPTVEYVSVYAAQNRNGESQPLVQTAV	1719	
QY	288	-----ALPEGGSGAFPPGHFKD-----	305	
DB	1720	TNIDRPKGLAFTDVDVDSIKIAWESPGQGVSRVRYTSSPDGIRELFPAPGDGDDTABL	1779	
QY	306	-----PKRLYCKN-----	313	
DB	1780	QGLRPGSEYTVSVVALHDMESQPLIGIOSTAIPAPTNLKLQVTPTSFTAQWIAPSVOL	1839	
QY	314	GGFFLRHIDPGRVDGVREKSDPHIKIQLOAERGGVSVIKGVCANRY	369	
DB	1840	TGVRVRVNP-----KEXTGPMKEINLSPDSSVIVSGLMVATKYEVSVYALKD--TL	1889	
QY	370	ASK-----CVTDEC-----PFPERLESNNY-----	389	
DB	1890	TGRPAQGVITLLENVSPRRARVTDATETITISMRKTETITGFOVDAIPAANGPTPVQR	1949	
QY	390	-----NYRSRKYT-----	398	
DB	1950	SISPDVRSYITIGLQGTGYDKIHLTYLNDNARSSPVIIDASTAIDAPSNLRFITTPNSL	2009	
QY	399	--SWYVALKRTGY-----KLASK-----TG--PGQKAILFL-----	426	
DB	2010	LVSQAPRARITGYIIKYEKPGSPREVVRPRPGVTEATITGLEPGTEYTVIVTALKNN	2069	
QY	427	----PM--SAASELPQLVTLPHNHLHGPEILDVPST	457	
DB	2070	QKSEPLIGRKKTDLPQLVTLPHNHLHGPEILDVPST	2106	
RESULT 5				
FINC_CHK				
ID	FINC_CHK	STANDARD;	PRT;	1256 AA.

AC	P11722; Q90921;
DT	01-OCT-1989 (Rel. 12, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	15-MAR-2004 (Rel. 43, Last annotation update)
DE	Fibronectin (FN) (Fragments).
GN	FN1.
OS	Gallus gallus (Chicken).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC	Gallus.
OX	NCBI_TaxID=9031;
RN	[1]
RP	SEQUENCE OF 1-50 FROM N.A.
RA	MEDLINE=83117850; PubMed=6572007;
RA	Hirano H., Yamada Y., Sullivan M., de Crombrughe B., Pastan I.,
RA	Yamada K.M.;
RT	"Isolation of genomic DNA clones spanning the entire fibronectin
RT	gene.";
RL	Proc. Natl. Acad. Sci. U.S.A. 80:46-50(1983).
RN	[2]
RP	SEQUENCE OF 51-1256 FROM N.A.
RC	STRAIN=White leghorn;
RA	Norton P.A.;
RN	Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
RN	[3]
RP	SEQUENCE OF 227-415 FROM N.A.
RX	MEDLINE=96183658; PubMed=8603103;
RA	Gehris A.L., Brandli D.W., Lewis S.D., Bennett V.D.;
RT	"The exon encoding the fibronectin type III-9 repeat is
RT	constitutively included in the mRNA from chick limb mesenchyme and
RT	cartilage.";
RL	Biochim. Biophys. Acta 1311:5-12(1996).
RN	[4]
RP	SEQUENCE OF 327-599 FROM N.A.
RX	MEDLINE=88050950; PubMed=2823899;
RA	Kubomura S., Obara M., Karasaki Y., Taniguchi H., Gotoh S.,
RA	Tsuda T., Higashi K., Ohsato K., Hiarno H.;
RT	"Genetic analysis of the cell binding domain region of the chicken
RT	fibronectin gene.";
RL	Biochim. Biophys. Acta 910:171-181(1987).
RN	[5]
RP	SEQUENCE OF 413-1256 FROM N.A.
RX	MEDLINE=88142820; PubMed=2830487;
RA	Norton P.A., Hynes R.O.;
RT	"Alternative splicing of chicken fibronectin in embryos and in normal
RT	and transformed cells.";
RL	Mol. Cell. Biol. 7:4297-4307(1987).
CC	-!- FUNCTION: Fibronectins bind cell surfaces and various compounds
CC	including collagen, fibrin, heparin, DNA, and actin. Fibronectins
CC	are involved in cell adhesion, cell motility, opsonization, wound
CC	healing, and maintenance of cell shape.
CC	-!- SUBUNIT: MOSTLY HETERODIMERS OR MULTIMERS OF ALTERNATIVELY SPLICED
CC	VARIANTS, CONNECTED BY 2 DISULFIDE BONDS NEAR THE CARBOXYL ENDS;
CC	-!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC	-!- ALTERNATIVE PRODUCTS:
CC	Event=Alternative splicing; Named isoforms=1;
CC	Comment=A number of isoforms are produced. Each of the "extra
CC	domain" and the connecting strand 3 are present in some forms of
CC	fibronectin and absent in others;
CC	Name=1;
CC	Isoid=P11722-1; Sequences=Displayed;
CC	-!- TISSUE SPECIFICITY: Plasma FN (soluble dimeric form) is secreted
CC	by hepatocytes. Cellular FN (dimeric or cross-linked multimeric
CC	forms), made by fibroblasts, epithelial and other cell types, is
CC	deposited as fibrils in the extracellular matrix.
CC	-!- PTM: Sulfated (By similarity).
CC	-!- SIMILARITY: Contains at least 2 fibronectin type I domains.
CC	-!- SIMILARITY: Contains at least 8 fibronectin type III domains.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; V00432; CAA23714.1; -  
 EMBL; U21327; AAA73566.1; -  
 EMBL; X06533; CAA29781.1; -  
 EMBL; M26186; AAA48772.1; ALT\_SEQ.  
 EMBL; U20386; AAB01062.1; -  
 PIR; A28512; A28512.  
 PIR; A29355; A29355.  
 PIR; S71465; S71465.  
 HSSP; P02751; 1FNH.  
 InterPro; IPR000083; Fibrnctnl.  
 InterPro; IPR008957; FN\_III-like.  
 InterPro; IPR003961; FN\_III.  
 InterPro; IPR000562; FN\_Type\_II.  
 InterPro; IPR003962; FNIII\_subd.  
 PRINTS; PR00014; FNTYPEIII.  
 SMART; SM00058; FN3; 3.  
 SMART; SM00060; FN3; 9.  
 PROSITE; PS00023; FIBRONECTIN\_2; PARTIAL.  
 PROSITE; PS01253; FIBRONECTIN\_1; 2.  
 Glycoprotein; Plasma; Heparin-binding; Acute phase; Cell adhesion;  
 Repeat; Sulfation; Alternative splicing.  
 KW NON TER 1 1  
 FT NON CONS 50 51  
 FT DOMAIN 236 509  
 FT DOMAIN 690 961  
 FT DOMAIN 1153 1226  
 FT DOMAIN 327 415  
 FT DOMAIN 416 509  
 FT DOMAIN 510 599  
 FT DOMAIN 600 689  
 FT DOMAIN 690 781  
 FT DOMAIN 782 871  
 FT DOMAIN 872 961  
 FT DOMAIN 962 1082  
 FT DOMAIN 1083 1152  
 FT DOMAIN 1174 1218  
 FT DOMAIN 1219 1256  
 FT SITE 493 495  
 FT DISULFID 1176 1205  
 FT DISULFID 1203 1215  
 FT DISULFID 1221 1248  
 FT CARBOHYD 122 122  
 FT CARBOHYD 1078 1078  
 FT CARBOHYD 1034 1034  
 FT CARBOHYD 1035 1035  
 FT CONFLICT 516 516  
 FT CONFLICT 569 572  
 FT NON TER 1256 1256  
 SQ SEQUENCE 1256 AA; 137435 MW; 345A4CA0E4D71D9B CRC64;  
 Query Match 50.5%; Score 1210.5; DB 1; Length 1256;  
 Best Local Similarity 76.3%; Pred. No. 9.7e-78;  
 Matches 238; Conservative 23; Mismatches 32; Indels 19; Gaps 2;  
 1 PDLRFNIGPDNRVTVWAPPSSIDLNFVLVRVSPVKNEEDVAELTSPSDNAVLTNLL 60  
 239 PDLRFNIGPDNRVTVWAPPSSIDLNFVLVRVSPVKNEEDVAELTSPSDNAVLTNLL 298  
 61 PGTEYVVSVEQHESTPLRGOKTGLDPSGTIDFSDITANSFTVHWIAPRATITGYR 120  
 299 PGTEYVVRVVAEQHESAPLSGQKTLGDSPTGLDPSDITANSFTVHWIAPRATITGYK 358  
 121 IRHPEHFGSPREDRVPHRSNITLTNLTPTGTEYVVSIVVALNGRESPLLIQOQSTVSD 180  
 359 IRHPEHGVGRPREDRVPPNSITLTNLTPTGTEYVVSIIAVNGRESVPLVGOQTIVSD 418  
 181 VPRDLVVAATPTSLLSISWDAPAVTVYRITYTGETCGNSPVQEFVTPGSKSTATISGLK 240

Db 419 VPRDLVNPSTPSLSISWDAPAVTVYRITYTGETCGNSPVQEFVTPGSKSTATISGLK 478  
 Qy 241 PGVDYITTVAVTGRGDSPASSKPIISYNTXTEIDKPSMA-----AGSITTLP----- 287  
 Db 479 PGVDYITTVAVTGRGDSPASSKPIISYNTXTEIDKPSMA-----AGSITTLP----- 538  
 Qy 288 -----ALPEDG 293  
 Db 539 TGYRVTVAVPKKG 550  
 RESULT 6  
 PISC\_PLEWA  
 ID PISC\_PLEWA STANDARD; PRT; 1328 AA.  
 AC Q91289;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Fibronectin (FN) (Fragment).  
 OS Pleurodeles waltli (Iberian ribbed newt).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae;  
 OC Pleurodeles  
 OC NCBI\_TaxID=8319;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94363379; PubMed=8081872;  
 RA Cavalier L., Riou J., Desimone D.W.;  
 RT "Amphibian Pleurodeles waltli fibronectin: cDNA cloning and  
 development expression of spliced variants.";  
 RL Cell Adhes. Commun. 1:83-91(1993).  
 CC -!- FUNCTION: Fibronectins bind cell surfaces and various compounds  
 including collagen, fibrin, heparin, DNA, and actin. Fibronectins  
 are involved in cell adhesion, cell motility, opsonization, wound  
 healing, and maintenance of cell shape (By similarity).  
 CC -!- SUBUNIT: DIMERS OR MULTIMERS OF ALTERNATIVELY SPLICED VARIANTS,  
 CONNECTED BY 2 DISULFIDE BONDS NEAR THE CARBOXYL ENDS (BY  
 SIMILARITY).  
 CC -!- SIMILARITY: Contains at least 3 fibronectin type I domains.  
 CC -!- SIMILARITY: Contains at least 10 fibronectin type III domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X56813; CAA47292.1; -  
 HSSP; P02751; 1FNH.  
 InterPro; IPR006209; EGF like.  
 InterPro; IPR000083; Fibrnctnl.  
 InterPro; IPR008957; FN\_III-like.  
 InterPro; IPR003961; FN\_III.  
 InterPro; IPR003962; FNIII\_subd.  
 Pfam; PF00039; fn1; 3.  
 Pfam; PF00041; fn3; 11.  
 PRINTS; PR00014; FNTYPEIII.  
 SMART; SM00058; FN1; 3.  
 SMART; SM00060; FN3; 9.  
 PROSITE; PS00022; EGF 1; 1.  
 PROSITE; PS01253; FIBRONECTIN\_1; 1.  
 Glycoprotein; Plasma; Heparin-binding; Acute phase; Cell adhesion;  
 Repeat.  
 KW NON TER 1 1  
 FT DNA BIND <1 13 BY SIMILARITY.  
 FT DOMAIN 203 477 CELL-ATTACHMENT.  
 FT DOMAIN 658 929 HEPARIN-BINDING 2 (BY SIMILARITY).  
 FT DOMAIN 1157 1288 HEPARIN-BINDING 2 (BY SIMILARITY).  
 FT DOMAIN <1 14 FIBRONECTIN TYPE-III.  
 FT DOMAIN 15 110 FIBRONECTIN TYPE-III.  
 FT DOMAIN 111 201 FIBRONECTIN TYPE-III (EXTRA DOMAIN).

```
FT DOMAIN 202 292 FIBRONECTIN TYPE-III.
FT DOMAIN 293 383 FIBRONECTIN TYPE-III.
FT DOMAIN 384 477 FIBRONECTIN TYPE-III.
FT DOMAIN 478 567 FIBRONECTIN TYPE-III.
FT DOMAIN 568 657 FIBRONECTIN TYPE-III.
FT DOMAIN 658 749 FIBRONECTIN TYPE-III.
FT DOMAIN 750 838 FIBRONECTIN TYPE-III.
FT DOMAIN 839 929 FIBRONECTIN TYPE-III.
FT DOMAIN 930 1064 CONNECTING STRAND 3 (CS-3) (V REGION).
FT DOMAIN 1065 1143 FIBRONECTIN TYPE-III.
FT DOMAIN 1155 1199 FIBRONECTIN TYPE-I.
FT DOMAIN 1200 1243 FIBRONECTIN TYPE-I.
FT DOMAIN 1245 1287 FIBRONECTIN TYPE-I.
FT DISULFID 1157 1186 BY SIMILARITY.
FT DISULFID 1184 1196 BY SIMILARITY.
FT DISULFID 1202 1229 BY SIMILARITY.
FT DISULFID 1227 1240 BY SIMILARITY.
FT DISULFID 1247 1270 BY SIMILARITY.
FT DISULFID 1268 1284 BY SIMILARITY.
FT DISULFID 1320 1320 INTERCHAIN (WITH C-1324) (BY SIMILARITY).
FT DISULFID 1324 1324 INTERCHAIN (WITH C-1320) (BY SIMILARITY).
FT SITE 461 463 CELL ATTACHMENT SITE.
FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1328 AA; 145037 MW; E31BF7968A1D1E74 CRC64;

Query Match 43.8%; Score 1051.5; DB 1; Length 1328;
Best Local Similarity 71.1%; Pred. No. 1.9e-66;
Matches 197; Conservative 41; Mismatches 38; Indels 1; Gaps 1;

QY 2 TDLRFTNIGDTRVTPWAPPSSIDLTLNVLRYSPVKNEDVAELSPSDNAVLTNLLP 61
Db 207 TDLRFTNVGDSMLVTSAPPSSVLSFLRVYVPSKNEEDAAELTSPSDNVMVLTNLLP 266
QY 62 GTEYVVSVSRYQHSTPLRGKQKGLDPTGIDFSDITANSFTVHWIAPRATITGYPI 121
Db 267 GTEYIVSVFAYVERESTPLTGVRGLDPTGIDFSDITSSFTVYVWAPRATVGYKI 326
QY 122 RHPEHFGSGPRDR-VPHSRNSITLTLNTPGTYVVSVALNGREESPLLICQOSTVSD 180
Db 327 QYHPETGGAGQKERCVPFSRNSLTTLNTPGTYVVSIFAVNGRESVPLVQQATVSD 386
QY 181 VPRDLEVVAAPTSLTSLISWDAPAVTVYRITYTGETGNSPVQEFVTPGSKSTATISGLK 240
Db 387 TPTNLEVTSTPTSMISMDAPPVGVYRITYTGETGTPVQEFVTPGDRSDAPIRGLK 446
QY 241 PGVDYITVYVAVTGRGDSPASSPISINRYTETDKPS 277
Db 447 PGAEYITVYVAVTGRGDSPASSKPVTVTHKTVVDKPT 483
```

## RESULT 7

```
FINC_XENLA STANDARD; PRT; 2481 AA.
AC Q91740;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Fibronectin precursor.
GN FN1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92111942; PubMed=1730390;
RA Desimone D.W., Norton P.A., Hynes R.O.;
RT "Identification and characterization of alternatively spliced
RL fbronectin mRNAs expressed in early Xenopus embryos.";
RL Dev. Biol. 149:357-369(1992).
CC -!- FUNCTION: Fibronectins bind cell surfaces and various compounds
CC including collagen, fibrin, heparin, DNA, and actin. Fibronectins
```

```
CC are involved in cell adhesion, cell motility, opsonization, wound
CC healing, and maintenance of cell shape (By similarity).
CC -!- SUBUNIT: DIMERS OR MULTIMERS OF ALTERNATIVELY SPLICED VARIANTS,
CC CONNECTED BY 2 DISULFIDE BONDS NEAR THE CARBOXYL ENDS (BY
CC SIMILARITY).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=1;
CC Comment=A number of isoforms are produced. Each of the "extra
CC domain" and the connecting strand 3 are present in some forms of
CC fibronectin and absent in others;
CC Name=1;
CC IsoId=Q91740-1; Sequence=Displayed; EMBRYO, CELLULAR FORMS OF
CC -!- TISSUE SPECIFICITY: IN EARLY XENOPUS EMBRYO, CELLULAR FORMS OF
CC FIBRONECTIN PREDOMINATE WHICH INCLUDE BOTH EXTRA DOMAINS. IN
CC FIBRONECTIN OF EMBRYONIC AND ADULT LIVER THE CONNECTING STRAND 3
CC CAN BE EITHER COMPLETELY EXCLUDED OR INCLUDED.
CC -!- SIMILARITY: Contains 12 fibronectin type I domains.
CC -!- SIMILARITY: Contains 2 fibronectin type II domains.
CC -!- SIMILARITY: Contains 17 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M77820; AAA49707.1; -.
CC HSSP; P02751; 2FN2.
CC InterPro; IPR006209; EGF like.
CC InterPro; IPR000083; Fibrinctn1.
CC InterPro; IPR008957; FN III-like.
CC InterPro; IPR003961; FN III.
CC InterPro; IPR000562; FN_Type_II.
CC Pfam; PF00039; fn1; 12.
CC Pfam; PF00040; fn2; 2.
CC Pfam; PF00041; fn3; 17.
CC PRINTS; PR00013; FNTYPEII.
CC ProDom; PD000995; FN_Type_II; 2.
CC SMART; SM00058; FN1; 12.
CC SMART; SM00059; FN2; 2.
CC SMART; SM00060; FN3; 17.
CC PROSITE; PS00022; EGF 1; 2.
CC PROSITE; PS01253; FIBRONECTIN 1; 11.
CC Glycoprotein; plasma; Heparin-binding; Acute phase; Cell adhesion;
CC Repeat; Alternative splicing; Signal
CC SIGNAL 1 31
CC FIBRONECTIN.
CC CHAIN 32 2481
CC FT DOMAIN 55 275 FIBRIN- AND HEPARIN-BINDING 1.
CC FT DOMAIN 309 609 COLLAGEN-BINDING.
CC FT DNA_BIND 907 1172 BY SIMILARITY.
CC FT DOMAIN 1358 1631 CELL-ATTACHMENT.
CC FT DOMAIN 1812 2082 HEPARIN-BINDING 2.
CC FT DOMAIN 2301 2432 FIBRIN-BINDING 2.
CC FT DOMAIN 53 93 FIBRONECTIN TYPE-I 1.
CC FT DOMAIN 98 141 FIBRONECTIN TYPE-I 2.
CC FT DOMAIN 142 185 FIBRONECTIN TYPE-I 3.
CC FT DOMAIN 187 231 FIBRONECTIN TYPE-I 4.
CC FT DOMAIN 232 276 FIBRONECTIN TYPE-I 5.
CC FT DOMAIN 307 346 FIBRONECTIN TYPE-I 6.
CC FT DOMAIN 346 405 FIBRONECTIN TYPE-II 1.
CC FT DOMAIN 406 470 FIBRONECTIN TYPE-II 2.
CC FT DOMAIN 469 512 FIBRONECTIN TYPE-I 7.
CC FT DOMAIN 517 559 FIBRONECTIN TYPE-I 8.
CC FT DOMAIN 560 603 FIBRONECTIN TYPE-I 9.
CC FT DOMAIN 610 707 FIBRONECTIN TYPE-III 1.
CC FT DOMAIN 708 809 FIBRONECTIN TYPE-III 2.
CC FT DOMAIN 810 904 FIBRONECTIN TYPE-III 3.
CC FT DOMAIN 905 995 FIBRONECTIN TYPE-III 4.
CC FT DOMAIN 996 1085 FIBRONECTIN TYPE-III 5.
CC FT DOMAIN 1086 1173 FIBRONECTIN TYPE-III 6.
CC FT DOMAIN 1174 1265 FIBRONECTIN TYPE-III 7.
```

FT	DOMAIN	1366	FIBRONECTIN TYPE-III 8 (EXTRA DOMAIN).
FT	DOMAIN	1357	FIBRONECTIN TYPE-III 9.
FT	DOMAIN	1447	FIBRONECTIN TYPE-III 10.
FT	DOMAIN	1448	FIBRONECTIN TYPE-III 11.
FT	DOMAIN	1538	FIBRONECTIN TYPE-III 12.
FT	DOMAIN	1632	FIBRONECTIN TYPE-III 13 (EXTRA DOMAIN).
FT	DOMAIN	1722	FIBRONECTIN TYPE-III 14.
FT	DOMAIN	1812	FIBRONECTIN TYPE-III 15.
FT	DOMAIN	1904	FIBRONECTIN TYPE-III 16.
FT	DOMAIN	1993	FIBRONECTIN TYPE-III 17.
FT	DOMAIN	2083	CONNECTING STRAND 3 (CS-3) (V REGION).
FT	DOMAIN	2206	FIBRONECTIN TYPE-I 10.
FT	DOMAIN	2299	FIBRONECTIN TYPE-I 11.
FT	DOMAIN	2344	FIBRONECTIN TYPE-I 12.
FT	DOMAIN	2388	CELL ATTACHMENT SITE.
FT	SITE	1615	BY SIMILARITY.
FT	DISULFID	55	BY SIMILARITY.
FT	DISULFID	79	BY SIMILARITY.
FT	DISULFID	90	BY SIMILARITY.
FT	DISULFID	100	BY SIMILARITY.
FT	DISULFID	128	BY SIMILARITY.
FT	DISULFID	146	BY SIMILARITY.
FT	DISULFID	170	BY SIMILARITY.
FT	DISULFID	189	BY SIMILARITY.
FT	DISULFID	216	BY SIMILARITY.
FT	DISULFID	234	BY SIMILARITY.
FT	DISULFID	261	BY SIMILARITY.
FT	DISULFID	309	BY SIMILARITY.
FT	DISULFID	334	BY SIMILARITY.
FT	DISULFID	361	BY SIMILARITY.
FT	DISULFID	375	BY SIMILARITY.
FT	DISULFID	421	BY SIMILARITY.
FT	DISULFID	471	BY SIMILARITY.
FT	DISULFID	497	BY SIMILARITY.
FT	DISULFID	519	BY SIMILARITY.
FT	DISULFID	544	BY SIMILARITY.
FT	DISULFID	562	BY SIMILARITY.
FT	DISULFID	588	BY SIMILARITY.
FT	DISULFID	2301	BY SIMILARITY.
FT	DISULFID	2328	BY SIMILARITY.
FT	DISULFID	2346	BY SIMILARITY.
FT	DISULFID	2371	BY SIMILARITY.
FT	DISULFID	2390	BY SIMILARITY.
FT	DISULFID	2412	BY SIMILARITY.
FT	DISULFID	2459	INTERCHAIN (WITH C-2463) (BY SIMILARITY).
FT	DISULFID	2463	INTERCHAIN (WITH C-2459) (BY SIMILARITY).
FT	CARBOHYD	431	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	529	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	543	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	877	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1244	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1291	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2202	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2481	AA; 272678 MW; 7E47DF4P6CE72C93 CRC64;
SQ	SEQUENCE	42.6%;	Score 1020.5; DB 1; Length 2481;
		Best Local Similarity	68.7%; Pred. No. 6.9e-64;
		Matches 195;	Conservative 44; Mismatches 44; Indels 1; Gaps 1;
Qy	1	PTDLRFTNIGDTRVMTWAPPSPDLTNFLVRYSPVKNEDVAELSPSPDNVLTNLL	60
Db	1361	PTNLRFTNIGDNRVMTWAPPSPDLTNFLVRYSPVKNEDVAELSPSPDNVLTNLL	1420
Qy	61	PGTEYVSVSVYQHESTPLRGQKTGLDPTGIDFSDITANSFTVHWIAPRATITGYR	120
Db	1421	PFTYLVSVSVYQHESTPLRGQKTGLDPTGIDFSDITANSFTVHWIAPRATITGYR	1480
Qy	121	IRHPEHFGCRPRDRVPHSRNSTLTNLTPTGTEYVSVSVYQHESTPLRGQKTGLDPTGIDFSDITANSFTVHWIAPRATITGYR	180
Db	1481	IRYQLESGACRPRDRVPHSRNSTLTNLTPTGTEYVSVSVYQHESTPLRGQKTGLDPTGIDFSDITANSFTVHWIAPRATITGYR	1540
Qy	181	VPDLVFAVPTSLLSISWADAVTVYRYITVGETCGNSPVQFTVPGSKSTATISGLK	240
Db	1541	VPTDLEVTSSSPNTLTISWADAVTVYRYITVGETCGNSPVQFTVPGSKSTATISGLK	1600

Qy	241	PGVDYITVYAVTGRGDSPASSKPSISYNYTEIDKP-SMAAGSI	283
Db	1601	PGVSYTITVYAVTGRGDSPASSKPSLTIIHKTDVDQPIDMAVTDI	1644
		RESULT 8	
ID	FGF2_HUMAN	STANDARD;	PRT; 155 AA.
AC	P09038;		
DT	01-NOV-1988 (Rel. 09, Created)		
DT	01-NOV-1988 (Rel. 09, Last sequence update)		
DT	10-OCT-2003 (Rel. 42, Last annotation update)		
DE	Heparin-binding growth factor 2 precursor (HBGF-2) (Basic fibroblast growth factor) (BFGF) (Prostatropin).		
GN	FGF2 OR FGF8.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RN	SEQUENCE FROM N.A.		
RX	MEDLINE=87053817; PubMed=3780670;		
RA	Abraham J.A., Whang J.L., Tumolo A., Mergia A., Friedman J.,		
RA	Gospodarowicz D., Fiddes J.C.;		
RT	"Human basic fibroblast growth factor: nucleotide sequence and genomic organization."		
RL	EMBO J. 5:2523-2528(1986).		
RN	[2]		
RN	SEQUENCE FROM N.A.		
RX	MEDLINE=87217066; PubMed=3472745;		
RA	Abraham J.A., Whang J.L., Tumolo A., Mergia A., Fiddes J.C.;		
RT	"Human basic fibroblast growth factor: nucleotide sequence, genomic organization, and expression in mammalian cells."		
RL	Cold Spring Harb. Symp. Quant. Biol. 51:657-668(1986).		
RN	[3]		
RN	SEQUENCE FROM N.A.		
RX	MEDLINE=87213238; PubMed=3579930;		
RA	Sommer A., Brewer M.T., Thompson R.C., Moscatelli D., Presta M.,		
RA	Rifkin D.B.;		
RT	"A form of human basic fibroblast growth factor with an extended amino terminus."		
RL	Biochem. Biophys. Res. Commun. 144:543-550(1987).		
RN	[4]		
RN	SEQUENCE FROM N.A.		
RX	MEDLINE=87162468; PubMed=2435575;		
RA	Kurokawa T., Sasada R., Iwane M., Igarashi K.;		
RT	"Cloning and expression of cDNA encoding human basic fibroblast growth factor."		
RL	FEBS Lett. 213:189-194(1987).		
RN	[5]		
RN	SEQUENCE FROM N.A.		
RX	MEDLINE=89184522; PubMed=2539817;		
RA	Prats H., Kaghad M., Prats A.C., Klagsbrun M., Lelias J.M.,		
RA	Liauzun P., Chalou P., Tauber J.P., Anallric F., Smith J.A.,		
RT	"High molecular mass forms of basic fibroblast growth factor are initiated by alternative CUG codons."		
RL	Proc. Natl. Acad. Sci. U.S.A. 86:1836-1840(1989).		
RN	[6]		
RN	SEQUENCE OF 10-35.		
RX	MEDLINE=86275260; PubMed=3732516;		
RA	Gautschi P., Frater-Schroeder M., Boehlen P.;		
RT	"Partial molecular characterization of endothelial cell mitogens from human brain: acidic and basic fibroblast growth factors."		
RL	FEBS Lett. 204:203-207(1986).		
RN	[7]		
RN	SEQUENCE OF 10-39.		
RX	MEDLINE=86186784; PubMed=3964259;		
RA	Gimenez-Gallego G., Conn G., Hatcher V.B., Thomas K.A.;		
RT	"Human brain-derived acidic and basic fibroblast growth factors: amino terminal sequences and specific mitogenic activities."		
RL	Biochem. Biophys. Res. Commun. 135:541-548(1986).		
RN	[8]		

SEQUENCE OF 2-22.  
RX MEDLINE=87156686; PubMed=2435284;  
RA Story M.T., Esch F., Shimazaki S., Sasse J., Jacobs S.C., Lawson R.K.;  
RT "Amino-terminal sequence of a large form of basic fibroblast growth  
factor isolated from human benign prostatic hyperplastic tissue.";  
RL Biochem. Biophys. Res. Commun. 142:702-709(1987).  
RN [9]  
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).  
RX MEDLINE=92121151; PubMed=1769963;  
RA Ago H., Kitagawa Y., Fujishima A., Matsuura Y., Katsube Y.;  
RT "Crystal structure of basic fibroblast growth factor at 1.6-A  
resolution.";  
RL J. Biochem. 110:360-363(1991).  
RN [10]  
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).  
RX MEDLINE=91195367; PubMed=1707542;  
RA Eriksson A.E., Cousins L.S., Weaver L.H., Matthews B.W.;  
RT "Three-dimensional structure of human basic fibroblast growth  
factor.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:3441-3445(1991).  
RN [11]  
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).  
RX MEDLINE=91195368; PubMed=1849658;  
RA Zhang J., Cousins L.S., Barr P.J., Sprang S.R.;  
RT "Three-dimensional structure of human basic fibroblast growth factor,  
a structural homolog of interleukin 1 beta.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:3446-3451(1991).  
RN [12]  
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).  
RX MEDLINE=91095983; PubMed=1702556;  
RA Zhu X., Komiya H., Chirino A., Faham S., Fox G.M., Arakawa T.,  
Hsu B.T., Rees D.C.;  
RT "Three-dimensional structures of acidic and basic fibroblast growth  
factors.";  
RL Science 251:90-93(1991).  
RN [13]  
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).  
RX MEDLINE=9400464; PubMed=7691311;  
RA Eriksson A.E., Cousins L.S., Matthews B.W.;  
RT "Refinement of the structure of human basic fibroblast growth factor  
at 1.6-A resolution and analysis of presumed heparin binding sites by  
selenate substitution.";  
RL Protein Sci. 2:1274-1284(1993).  
RN [14]  
RP STRUCTURE BY NMR.  
RX MEDLINE=97040521; PubMed=8885834;  
RA Moy F.J., Seddon A.P., Boehlen P., Powers R.;  
RT "High-resolution solution structure of basic fibroblast growth factor  
determined by multidimensional heteronuclear magnetic resonance  
spectroscopy.";  
RL Biochemistry 35:13552-13561(1996).  
CC -!- FUNCTION: The heparin-binding growth factors are angiogenic agents  
in vivo and are potent mitogens for a variety of cell types in  
vitro. There are differences in the tissue distribution and  
concentration of these 2 growth factors.  
CC -!- SUBUNIT: Monomer.  
CC -!- MISCELLANEOUS: This protein binds heparin more strongly than does  
AFGF.  
CC -!- SIMILARITY: Belongs to the heparin-binding growth factors family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M17599; AAA52534.1; ALT\_INIT.  
DR EMBL; X04431; CAA28027.1; -.  
DR EMBL; X04432; CAA28028.1; -.  
DR EMBL; X04433; CAA28029.1; -.  
DR EMBL; M27968; AAA52448.1; -.

DR EMBL; J04513; AAA52533.1; ALT\_INIT.  
DR PIR; A32398; A32398.  
DR PDB; 2FGF; 15-APR-92.  
DR PDB; 4FGF; 15-JUL-93.  
DR PDB; 1FGA; 15-JUL-93.  
DR PDB; 1BFB; 03-APR-96.  
DR PDB; 1BFC; 03-APR-96.  
DR PDB; 1BFF; 16-JUN-97.  
DR PDB; 1BFG; 31-JAN-94.  
DR PDB; 2BFH; 30-APR-94.  
DR PDB; 1BLA; 08-NOV-96.  
DR PDB; 1BLD; 08-NOV-96.  
DR PDB; 1BAS; 31-OCT-93.  
DR PDB; 1CVS; 28-JAN-00.  
DR PDB; 1EV2; 31-MAY-00.  
DR PDB; 1FQ9; 27-SEP-00.  
DR PDB; 1114; 20-JUN-01.  
DR PDB; 111L; 20-JUN-01.  
DR GENE; HGNC:3676; FGF2.  
DR MIM; 134920; -.  
DR GO; GO:0005737; C:cytoplasm; TAS.  
DR GO; GO:0005615; C:extracellular space; TAS.  
DR GO; GO:0000187; P:activation of MAPK; TAS.  
DR GO; GO:0006935; P:chemotaxis; TAS.  
DR GO; GO:0007397; P:histogenesis and organogenesis; TAS.  
DR GO; GO:0008284; P:positive regulation of cell proliferation; TAS.  
DR GO; GO:0007265; P:RAS protein signal transduction; TAS.  
DR InterPro; IPR008996; Cytok\_IL1\_like.  
DR InterPro; IPR002348; IL1\_HBGF.  
DR Pfam; PF00167; FGF; 1.  
DR PRINTS; PR00262; IL1\_HBGF.  
DR ProDom; PD000831; IL1\_HBGF; 1.  
DR SMART; SM00443; FGF; 1.  
DR PROSITE; PS00247; HBGF\_FGF; 1.  
KW Growth factor; Mitogen; Angiogenesis; Heparin-binding;  
KW 3D-structure.  
FT PROPEP 1 9  
FT CHAIN 10 155 HEPARIN-BINDING GROWTH FACTOR 2.  
FT SITE 46 48 CELL ATTACHMENT SITE (POTENTIAL).  
FT SITE 88 90 CELL ATTACHMENT SITE (POTENTIAL).  
FT BINDING 27 31 HEPARIN (POTENTIAL).  
FT BINDING 116 119 HEPARIN (POTENTIAL).  
FT STRAND 30 34 HEPARIN (POTENTIAL).  
FT TURN 35 38  
FT STRAND 39 43  
FT TURN 45 46  
FT STRAND 49 52  
FT TURN 55 56  
FT TURN 58 60  
FT STRAND 62 66  
FT TURN 69 70  
FT STRAND 71 76  
FT TURN 77 80  
FT STRAND 81 85  
FT TURN 87 88  
FT STRAND 91 94  
FT HELIX 99 101  
FT STRAND 103 107  
FT TURN 109 110  
FT STRAND 113 117  
Query Match 34.2%; Score 820; DB 1; Length 155;  
Best Local Similarity 99.4%; Pred. No. 2.2e-51;  
Matches 154; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 278 MAAGSITLPAIPEDGGSGAPPPGHFKDPKRLYCKNGGFFLRIHPDGRVDGVREKSDPHI 337  
DB 1 MAAGSITLPAIPEDGGSGAPPPGHFKDPKRLYCKNGGFFLRIHPDGRVDGVREKSDPHI 60  
QY 338 KIQLOAERGVSIVKVCANRYLAMKEDGRLLASKCVTDECFFERLESNNYTSRKY 397  
DB 61 KIQLOAERGVSIVKVCANRYLAMKEDGRLLASKCVTDECFFERLESNNYTSRKY 120





```
QY 398 TSWYVALKRTGYKLGSKTGPQKAILFLPMSAAS 432
:|||||
Db 121 SSWYVALKRTGYKLGSKTGPQKAILFLPMSAKS 155

RESULT 10
FGF2_SHEEP
ID FGF2_SHEEP STANDARD; PRT; 155 AA.
AC P20003;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heparin-binding growth factor 2 precursor (HBGF-2) (Basic fibroblast
DE growth factor) (BFGF) (Prostatropin).
GN FGF2 OR FGF-2.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RA Sutton R., Ward W.G., Raphael K.A., Cam G.R.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 9-155.
RX MEDLINE=8055577; PubMed=3678486;
RA Simpson R.J., Moritz R.L., Lloyd C.J., Fabri L.J., Nice E.C.,
RA Rubira M.R., Burgees A.W.;
RT "Primary structure of ovine pituitary basic fibroblast growth
RT factor.";
RT FEBS Lett. 224:128-132(1987).
CC -!- FUNCTION: The heparin-binding growth factors are angiogenic agents
CC in vivo and are potent mitogens for a variety of cell types in
CC vitro. There are differences in the tissue distribution and
CC concentration of these 2 growth factors.
CC -!- SUBUNIT: Monomer.
CC -!- MISCELLANEOUS: This protein binds heparin more strongly than does
CC aFGF.
CC -!- SIMILARITY: Belongs to the heparin-binding growth factors family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L36136; AAA31519.1; -.
DR HSP; P09038; 1BFF.
DR InterPro; IPR008996; Cytok IL1-like.
DR InterPro; IPR002348; IL1_HBGF.
DR Pfam; PF00167; FGF; 1.
DR PRINTS; PD00262; IL1HBGF.
DR PRODOM; PD000831; IL1_HBGF.
DR SMART; SM00442; FGF; 1.
DR PROSITE; PS00247; HBGF_FGF; 1.
KW Growth factor; Mitogen; Angiogenesis; Heparin-binding.
FT PROPEP 1 9
FT CHAIN 10 155 HEPARIN-BINDING GROWTH FACTOR 2.
FT SITE 45 48 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 87 90 CELL ATTACHMENT SITE (POTENTIAL).
FT BINDING 27 31 HEPARIN (POTENTIAL).
FT BINDING 116 119 HEPARIN (POTENTIAL).
FT BINDING 155 AA; 17280 MW; B5F2364BA610606D CRC64;
SQ
Query Match 33.6%; Score 805; DB 1; Length 155;
Best Local Similarity 97.4%; Pred. No. 2.6e-50;
Matches 151; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 278 MAAGSITTLPALPEDGSGSAPFPQKRLYCKNGGFFLRIHPDGRVDGVREKSDPHI 337
:|||||
Db 1 MAAGSITTLPALPEDGSGSAPFPQKRLYCKNGGFFLRIHPDGRVDGVREKSDPHI 60
338 KLQQAEEGVVSIKGVCANRYLAMKEDGRLLASKCVTDECFEERLESNNYNTYRSKY 397
|||||
Db 61 KLQQAEEGVVSIKGVCANRYLAMKEDGRLLASKCVTDECFEERLESNNYNTYRSKY 120
398 TSWYVALKRTGYKLGSKTGPQKAILFLPMSAAS 432
:|||||
Db 121 SSWYVALKRTGYKLGSKTGPQKAILFLPMSAKS 155

RESULT 11
FGF2_RAT
ID FGF2_RAT STANDARD; PRT; 154 AA.
AC P13109;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heparin-binding growth factor 2 precursor (HBGF-2) (Basic fibroblast
DE growth factor) (BFGF) (Prostatropin).
GN FGF2 OR FGF-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Ovary;
RX MEDLINE=89061721; PubMed=3196337;
RA Shimasaki S., Emoto N., Koba A., Mercado M., Shibata F.,
RA Cooksey K., Baird A., Ling N.;
RT "Complementary DNA cloning and sequencing of rat ovarian basic
RT fibroblast growth factor and tissue distribution study of its mRNA.";
RL Biochem. Biophys. Res. Commun. 157:256-263(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=88262516; PubMed=3387229;
RA Kurokawa T., Seno M., Igarashi K.;
RT "Nucleotide sequence of rat basic fibroblast growth factor cDNA.";
RL Nucleic Acids Res. 16:5201-5201(1988).
RN [3]
RP SEQUENCE OF 1-28 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Testis;
RX MEDLINE=97200905; PubMed=9048734;
RA Pasumarthi K.B.S., Jin Y., Cattini P.A.;
RT "Cloning of the rat fibroblast growth factor-2 promoter region and
RT its response to mitogenic stimuli in glioma C6 cells.";
RL J. Neurochem. 68:898-908(1997).
RN [4]
RP SEQUENCE OF 35-154 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=92329546; PubMed=1378302;
RA El-Husseini A.E.-D., Paterson J.A., Myal Y., Shiu R.P.C.;
RT "PCR detection of the rat brain basic fibroblast growth factor (bFGF)
RT mRNA containing a unique 3' untranslated region.";
RL Biochim. Biophys. Acta 1131:314-316(1992).
CC -!- FUNCTION: The heparin-binding growth factors are angiogenic agents
CC in vivo and are potent mitogens for a variety of cell types in
CC vitro. There are differences in the tissue distribution and
CC concentration of these 2 growth factors.
CC -!- SUBUNIT: Monomer.
CC -!- TISSUE SPECIFICITY: Found in all tissues examined.
CC -!- MISCELLANEOUS: This protein binds heparin more strongly than does
CC aFGF.
CC -!- SIMILARITY: Belongs to the heparin-binding growth factors family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC
```



```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M22427; AAA41210.1; -.
DR EMBL; U70285; CAA30265.1; -.
DR EMBL; U78079; AAC53225.1; -.
DR EMBL; X61697; CAA43863.1; -.
DR PIR; A31674; A31674.
DR HSP; P09038; 1BFF.
DR InterPro; IPR008996; Cytok IL1_like.
DR PRINTS; PR00167; FGF; 1.
DR PRODOM; PD000831; IL1_HBGF.
DR SMART; SM00442; FGF; 1.
DR PROSITE; PS00247; HBGF_FGF; 1.
DR Growth factor; Mitogen; Angiogenesis; Heparin-binding.
KW PROPEP 1
FT CHAIN 10 154 HEPARIN-BINDING GROWTH FACTOR 2.
FT BINDING 26 30 HEPARIN (POTENTIAL).
FT BINDING 115 118 HEPARIN (POTENTIAL).
SQ SEQUENCE 154 AA; 17139 MW; 1A0F14FF423D8403 CRC64;

Query Match 33.0%; Score 792.5; DB 1; Length 154;
Best Local Similarity 96.1%; Pred. No. 1.9e-49;
Matches 149; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 278 MAAGSITTLPALPEDGGGAPPPGHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHI 337
DB 1 MAAGSITSLPALPEDGG- GAFFPGHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHV 59

QY 338 KLOLQAEERGVSIGVCANRYLAMKEDGRLASKVCVTDECFFERLESNNYNTYRSKY 397
DB 60 KLOLQAEERGVSIGVCANRYLAMKEDGRLASKVCVTDECFFERLESNNYNTYRSKY 119

QY 398 TSMYVALKRTGYKLGSKTGPQKAILFLPMSAAS 432
DB 120 SSMYVALKRTGYKLGSKTGPQKAILFLPMSAKS 154

RESULT 12
FGF2 MOUSE
ID_FGF2_MOUSE STANDARD; PRT; 154 AA.
AC P15655;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heparin-binding growth factor 2 precursor (HBGF-2) (Basic fibroblast
DE growth factor) (BFGF) (Prostatropin).
GN FGF2 OR FGF-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90201563; PubMed=23118343;
RA Hebert J.M., Basiglio C., Goldfarb M., Haub O., Martin G.R.;
RT "Isolation of cDNAs encoding four mouse FGF family members and
RT characterization of their expression patterns during embryogenesis.";
RL Dev. Biol. 138:454-463 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J, A/J, and NOD/LtJ; TISSUE=Spleen;
RA Ma R.Z., Teuscher C.;
CC Submitted (May-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: The heparin-binding growth factors are angiogenic agents
CC in vivo and are potent mitogens for a variety of cell types in
CC vitro. There are differences in the tissue distribution and
CC concentration of these 2 growth factors.
CC -!- SUBUNIT: Monomer.
CC -!- MISCELLANEOUS: This protein binds heparin more strongly than does
CC aFGF.
CC -!- SIMILARITY: Belongs to the heparin-binding growth factors family.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M30644; AAA37621.1; -.
DR EMBL; AF065903; AAC17503.1; -.
DR EMBL; AF065904; AAC17504.1; -.
DR EMBL; AF065905; AAC17505.1; -.
DR PIR; C37360; C37360.
DR HSP; P09038; 1BFF.
DR MGD; MGI:95516; Fgf2.
DR GO; GO:0005615; C:extracellular space; IDA.
DR GO; GO:0045597; P:positive regulation of cell differentiation; IDA.
DR InterPro; IPR008996; Cytok_IL1_like.
DR PRINTS; PR00167; FGF; 1.
DR PRINTS; PR00262; IL1HBGF.
DR PRODOM; PD000831; IL1_HBGF; 1.
DR SMART; SM00442; FGF; 1.
DR PROSITE; PS00247; HBGF_FGF; 1.
DR Growth factor; Mitogen; Angiogenesis; Heparin-binding.
KW PROPEP 1
FT CHAIN 10 154 HEPARIN-BINDING GROWTH FACTOR 2.
FT BINDING 26 30 HEPARIN (POTENTIAL).
FT BINDING 115 118 HEPARIN (POTENTIAL).
SQ SEQUENCE 154 AA; 17153 MW; 689F677416274388 CRC64;

Query Match 32.4%; Score 777.5; DB 1; Length 154;
Best Local Similarity 94.2%; Pred. No. 2.2e-48;
Matches 146; Conservative 5; Mismatches 3; Indels 1; Gaps 1;

QY 278 MAAGSITTLPALPEDGGGAPPPGHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHI 337
DB 1 MAAGSITSLPALPEDGGA- AFGPHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHV 59

QY 338 KLOLQAEERGVSIGVCANRYLAMKEDGRLASKVCVTDECFFERLESNNYNTYRSKY 397
DB 60 KLOLQAEERGVSIGVCANRYLAMKEDGRLASKVCVTDECFFERLESNNYNTYRSKY 119

QY 398 TSMYVALKRTGYKLGSKTGPQKAILFLPMSAAS 432
DB 120 SSMYVALKRTGYKLGSKTGPQKAILFLPMSAKS 154

RESULT 13
FGF2 MONDO
ID_FGF2_MONDO STANDARD; PRT; 156 AA.
AC P48758;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heparin-binding growth factor 2 precursor (HBGF-2) (Basic fibroblast
DE growth factor) (BFGF) (Prostatropin).
GN FGF2.
OS Monodelphis domestica (Short-tailed grey opossum).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Didelphimorphia; Didelphidae; Monodelphis.
OX NCBI_TaxID=13616;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RX MEDLINE=94296558; PubMed=8024698;
RA Kusevitt D.P., Sabourin C.L.K., Sherburn T.E., Ley R.D.;
RT "Characterization of cDNA encoding basic fibroblast growth factor of
RT the marsupial Monodelphis domestica.";
RL DNA Cell Biol. 13:549-554 (1994).
CC -!- FUNCTION: The heparin-binding growth factors are angiogenic agents
CC in vivo and are potent mitogens for a variety of cell types in

```

```
CC      vitro. There are differences in the tissue distribution and
CC      concentration of these 2 growth factors.
CC      -!- SUBUNIT: Monomer.
CC      -!- MISCELLANEOUS: This protein binds heparin more strongly than does
CC      aFGF.
CC      -!- SIMILARITY: Belongs to the heparin-binding growth factors family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; Z15154; CAA78854.1; ALT_INIT.
CC      HSP; P09038; 1BFF.
CC      InterPro; IPR008996; Cytok_IL1_like.
CC      InterPro; IPR002348; IL1_HBGF.
CC      Pfam; PF00167; FGF; 1.
CC      PRINTS; PR00262; IL1HBGF.
CC      ProDom; PD000831; IL1_HBGF; 1.
CC      SMART; SM00442; FGF; 1.
CC      PROSITE; PS00247; HBGF_FGF; 1.
CC      Growth factor; Mitogen; Angiogenesis; Heparin-binding.
CC      PROPEP 1 9
CC      CHAIN 10 156 HEPARIN-BINDING GROWTH FACTOR 2.
CC      BINDING 28 32 HEPARIN (POTENTIAL).
CC      BINDING 117 120 HEPARIN (POTENTIAL).
CC      SEQUENCE 156 AA; 17303 MW; 7E655FCC49BF1209 CRC64;
CC      -----
CC      Query Match 31.5%; Score 754.5; DB 1; Length 156;
CC      Best Local Similarity 92.3%; Pred. No. 9.7e-47;
CC      Matches 144; Conservative 5; Mismatches 6; Indels 1; Gaps 1;
CC      -----
QY      278 MAAGSITTLPALPED-GGSGAFPFGHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPH 336
Db      |||||
Db      1 MAAGSITTLPALSDGGGGGAFPPGHPKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPH 60
QY      337 IKLQAEERGVSIVKGVCANRYLAKMKGRLKSCVTDECFERLESNNYTYRSRK 396
Db      |||||
Db      61 IKLQAEERGVSIVKGVCANRYLAKMKGRLKSCVTDECFERLESNNYTYRSRK 120
QY      397 YTSWYVALKRTGYKLGSKTGPQKAILFLPMSAAS 432
Db      |||||
Db      121 YSNWYVALKRTGYKLGSKTGPQKAILFLPMSAKS 156
CC      -----
RESULT 14
FGF2_CHICK STANDARD; PRT; 158 AA.
AC P48800;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heparin-binding growth factor 2 precursor (HBGF-2) (Basic fibroblast
DE growth factor) (BFGF).
GN FGF2 OR FGF-2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]_
RP SEQUENCE FROM N.A.
RX MEDLINE=93246053; PubMed=7683281;
RA Borja A.Z., Zeller R., Weijers C.;
RT "Expression of alternatively spliced bFGF first coding exons and
RT antisense mRNAs during chicken embryogenesis.";
RL Dev. Biol. 157:110-118(1993).
CC      -!- FUNCTION: The heparin-binding growth factors are angiogenic agents
CC      in vivo and are potent mitogens for a variety of cell types in
CC      vitro. There are differences in the tissue distribution and
```

```
CC      concentration of these 2 growth factors.
CC      -!- SUBUNIT: Monomer.
CC      -!- MISCELLANEOUS: This protein binds heparin more strongly than does
CC      aFGF.
CC      -!- SIMILARITY: Belongs to the heparin-binding growth factors family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; M95707; AAA48617.1; -.
CC      HSP; P09038; 1BFF.
CC      InterPro; IPR008996; Cytok_IL1_like.
CC      InterPro; IPR002348; IL1_HBGF.
CC      Pfam; PF00167; FGF; 1.
CC      PRINTS; PR00262; IL1HBGF.
CC      ProDom; PD000831; IL1_HBGF; 1.
CC      SMART; SM00442; FGF; 1.
CC      PROSITE; PS00247; HBGF_FGF; 1.
CC      Growth factor; Mitogen; Angiogenesis; Heparin-binding.
CC      PROPEP 1 12
CC      CHAIN 13 158 HEPARIN-BINDING GROWTH FACTOR 2.
CC      BINDING 30 34 HEPARIN (POTENTIAL).
CC      BINDING 119 122 HEPARIN (POTENTIAL).
CC      SEQUENCE 158 AA; 17374 MW; 7B69B684C17F1816 CRC64;
CC      -----
CC      Query Match 31.4%; Score 753; DB 1; Length 158;
CC      Best Local Similarity 91.6%; Pred. No. 1.3e-46;
CC      Matches 141; Conservative 5; Mismatches 8; Indels 0; Gaps 0;
CC      -----
QY      279 AAGSITTLPALPEDGGSGAFPFGHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHK 338
Db      |||||
Db      5 AAGSITTLPALPDGGGGGAFPPGHPKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHK 64
QY      339 IQLQAEERGVSIVKGVCANRYLAKMKGRLKSCVTDECFERLESNNYTYRSKYT 398
Db      |||||
Db      65 IQLQAEERGVSIVKGVCANRYLAKMKGRLKSCVTDECFERLESNNYTYRSKYS 124
QY      399 SWYVALKRTGYKLGSKTGPQKAILFLPMSAAS 432
Db      |||||
Db      125 DWYVALKRTGYKGPCKTGPQKAILFLPMSAKS 158
CC      -----
RESULT 15
FGF2_RABIT STANDARD; PRT; 137 AA.
AC P48759;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heparin-binding growth factor 2 (HBGF-2) (Basic fibroblast growth
DE factor) (BFGF) (Prostatropin) (Fragment).
GN FGF2.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand white; TISSUE=Smooth muscle;
RX MEDLINE=93343209; PubMed=8342599;
RA Winkles J.A., Friesel R., Alberts G.F., Janat M.F., Liau G.;
RT "Elevated expression of basic fibroblast growth factor in an
RT immortalized rabbit smooth muscle cell line.";
RL Am. J. Pathol. 143:518-527(1993).
CC      -!- FUNCTION: The heparin-binding growth factors are angiogenic agents
CC      in vivo and are potent mitogens for a variety of cell types in
CC      vitro. There are differences in the tissue distribution and
CC      concentration of these 2 growth factors.
```

```
CC -!- SUBUNIT: Monomer.
CC -!- MISCELLANEOUS: This protein binds heparin more strongly than does
CC aFGF.
CC
CC -!- SIMILARITY: Belongs to the heparin-binding growth factors family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L12034; AAA31248.1; -.
CC PIR; I46711; I46711.
CC HSSP; P09038; 1BFF.
CC InterPro; IPR008996; Cytok IL1 like.
CC InterPro; IPR002348; IL1_HBGF.
CC Pfam; PF00167; FGF; 1.
CC PRINTS; PR00262; IL1HBGF.
CC PRODOM; PD000831; IL1_HBGF; 1.
CC SMART; SM00442; FGF; 1.
CC PROSITE; PS00247; HBGF_FGF; 1.
CC Growth factor; Mitogen; Angiogenesis; Heparin-binding.
KW BINDING 18 22 HEPARIN (POTENTIAL).
FT BINDING 107 110 HEPARIN (POTENTIAL).
FT NON TER 137 137
SQ SEQUENCE 137 AA; 15418 MW; 0D9EE457B88E8C51 CRC64;

Query Match 30.8%; Score 738; DB 1; Length 137;
Best Local Similarity 99.3%; Pred. No. 1.2e-45;
Matches 136; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 287 PALPEDGGSGAFPPGHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLQLQAEER 346
Db 1 PALPEDGGSGAFPPGHFKDPKRLYCKNGGFFLRHPDGRVDGVREKSDPHIKLQLQAEER 60
Qy 347 GWSIKGVCANRYLAMKEDGRLASKCVTDECFFERLESNNYNTYRSKYTSWYVALKR 406
Db 61 GWSIKGVCANRYLAMKEDGRLASKCVTDECFFERLESNNYNTYRSKYTSWYVALKR 120
Qy 407 TGQYKLGSKTGPQKAI 423
Db 121 TGQYKLGSKTGPQKAI 137
```

Search completed: May 3, 2004, 13:07:49  
Job time : 16.9254 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 13:05:48 ; Search time 51.1726 Seconds  
(without alignments)

2817.753 Million cell updates/sec

Title: US-09-775-964-5

Perfect score: 2398

Sequence: 1 PTDLRRTNIGPDTRVTWAP.....LVTLPHNLRGPIIDVPST 457

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTRMBL\_25.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mmc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	824.5	34.4	196	4 P78443	P78443 homo sapien
2	762	31.8	153	11 Q925A3	Q925A3 mus musculus
3	760.5	31.7	2478	13 Q93406	Q93406 brachydanio
4	740.5	30.9	170	11 Q60487	Q60487 cavia porce
5	698	29.1	155	13 Q90Y92	Q90Y92 cynops pyr
6	676	28.2	130	6 Q77767	Q77767 canis famil
7	618	25.8	155	13 Q8QFR9	Q8qfr9 fugu rubrip
8	607.5	25.3	154	13 Q7ZKX5	Q7zdk5 brachydanio
9	585	24.4	111	6 Q9BDX1	Q9bdx1 macaca mula
10	561	23.4	108	6 Q9NLS7	Q9nls7 capreolus c
11	561	23.4	125	13 Q98TD8	Q98td8 cynops pyr
12	533	22.2	105	11 Q7TPG9	Q7tpg9 mus musculus
13	508	21.2	96	6 Q7YEN5	Q7yrn5 sus scrofa
14	484	20.2	109	11 Q925A1	Q925a1 mus musculus
15	480	20.0	112	11 Q925A2	Q925a2 mus musculus
16	476	19.8	101	13 P79706	P79706 cynops pyr

17	474.5	19.8	146	13	Q07659	Q07659 gallus gall
18	457	19.1	87	6	Q8WMP4	Q8wmp4 equus cabal
19	428.5	17.9	379	4	O95617	O95617 homo sapien
20	417	17.4	293	6	Q9XSG0	Q9xsg0 oryctolagus
21	415.5	17.3	810	11	Q8R3F3	Q8r3f3 mus musculus
22	396	16.5	1358	11	Q8BYI9	Q8byi9 mus musculus
23	394	16.4	1358	4	Q92752	Q92752 homo sapien
24	394	16.4	1358	4	Q15568	Q15568 homo sapien
25	383.5	16.0	354	4	Q9UQ56	Q9uq56 homo sapien
26	378.5	15.8	1356	11	Q05546	Q05546 rattus norv
27	358	14.9	1350	13	Q7T3T6	Q7t3t6 brachydanio
28	358	14.9	1353	13	Q00546	Q00546 gallus gall
29	341	14.2	76	6	Q9NOV2	Q9nov2 ovis aries
30	336	14.0	1532	13	Q90994	Q90994 gallus gall
31	332.5	13.9	114	4	Q16443	Q16443 homo sapien
32	332.5	13.9	114	4	O00527	O00527 homo sapien
33	329	13.7	1810	13	Q90824	Q90824 gallus gall
34	327	13.6	2019	11	Q64706	Q64706 mus musculus
35	327	13.6	2019	11	Q80YX2	Q80yx2 mus musculus
36	327	13.6	2110	11	Q80YX1	Q80yx1 mus musculus
37	326.5	13.6	614	13	Q90484	Q90484 brachydanio
38	318	13.3	619	11	Q62701	Q62701 rattus norv
39	317	13.2	1714	13	Q90995	Q90995 gallus gall
40	312.5	13.0	922	13	Q93405	Q93405 brachydanio
41	312	13.0	920	11	Q80UI4	Q80ui4 mus musculus
42	307.5	12.8	215	13	Q7TOK8	Q7tok8 ambystoma m
43	306.5	12.8	347	6	Q95KV5	Q95kv5 bos taurus
44	305	12.7	68	6	Q28692	Q28692 oryctolagus
45	302.5	12.6	4135	6	O18977	O18977 bos taurus

#### ALIGNMENTS

RESULT 1

P78443 ID P78443 PRELIMINARY; PRT; 196 AA.  
AC P78443;  
DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE 21 kDa basic fibroblast growth factor (BFGF).  
GN BFGF.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89184522; PubMed=2538817;  
RA Prats H., Kaghad M., Prats A.C., Klagsbrun M., Lelias J.M.,  
RA Liaunus P., Chalon P., Tauber J.P., Analtic F., Smith J.A., Caput D.;  
RT "High molecular mass forms of basic fibroblast growth factor are  
RT initiated by alternative CUG codons.";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:1836-1840(1989).  
RN [2]  
RP SEQUENCE OF 81-168 FROM N.A.  
RX MEDLINE=93038590; PubMed=1417798;  
RA Watson R., Anthony F., Pickett M., Lambden P., Masson G.M.,  
RA Thomas E.J.;  
RT "Reverse transcription with nested polymerase chain reaction shows  
RT expression of basic fibroblast growth factor transcripts in human  
RT granuloosa and cumulus cells from in vitro fertilisation patients.";  
RL Biochem. Biophys. Res. Commun. 187:1227-1231(1992).  
DR EMBL; J04513; AA52532.1; -;  
DR EMBL; S47380; AAD13853.1; -;  
DR HSSP; P09038; 1BFF.  
DR GO; GO:0008083; F: growth factor activity; IEA.  
DR InterPro; IPR008996; Cytok IL1\_like.  
DR InterPro; IPR002348; IL1\_HBGF.  
DR Pfam; PF00167; FGF; 1.  
DR PRINTS; PR00262; IL1HBGF.  
DR ProDom; PD000831; IL1\_HBGF; 1.





DT 01-NOV-1998 (TremBLrel. 08, Created)  
DT 01-NOV-1998 (TremBLrel. 08, Last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)  
DE Basic fibroblast growth factor (Fragment).  
GN BFGF.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Adrenal gland;  
RA Trochta O.A., Jacobs R.M., LaMarre J.;  
RT "The role bFGF in canine Hemangiosarcoma";  
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF060562; AAC35912.1; -.  
DR GO; GO:0008083; F:growth factor activity; IEA.  
DR InterPro; IPR008996; Cytok IL1\_like.  
DR InterPro; IPR002348; IL1\_HBGF.  
DR Pfam; PF00167; FGF; 1.  
DR PRINTS; PR00262; IL1HBGF.  
DR ProDom; PD000831; IL1\_HBGF; 1.  
DR SMART; SM00442; FGF; 1.  
DR PROSITE; PS00247; HBGF\_FGF; 1.  
FT NON\_TER 1  
FT NON\_TER 130  
SQ SEQUENCE 130 AA; 14902 MW; 21900876E878FAEA CRC64;  
Query Match 28.2%; Score 676; DB 6; Length 130;  
Best Local Similarity 96.9%; Pred. No. 4.2e-44;  
Matches 126; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 303 EKDPKRLKCKNGGFFLRIHPDGRVGVREKSPDHKILQQAERGVSIGKVCANRYLAM 362  
DB 1 FKPDKRLKCKNGGFFLRIHPDGRVGVREKSPDHKILQQAERGVSIGKVCANRYLAM 60  
QY 363 KEDGRLLASKCVTDECFPERLESNNYTYRSRKYTSWYVALKRTGQYKLGSKTGPQOKA 422  
DB 61 KEDGRLLASKCVTDECFPERLESNNYTYRSRKYTSWYVALKRTGQYKLGSKTGPQOKA 120  
QY 423 ILFLPMSAAS 432  
DB 121 ILFLPMSAKS 130  
RESULT 7  
Q8QFR9 PRELIMINARY; PRT; 155 AA.  
ID Q8QFR9  
AC Q8QFR9  
DT 01-JUN-2002 (TremBLrel. 21, Created)  
DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)  
DE Basic fibroblast growth factor.  
GN FGF2.  
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;  
OC Tetraodontidae; Tetraodontidae; Takifugu.  
OX NCBI\_TaxID=31033;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC Botcherby M.R.;  
RT "Comparative vertebrate genomic sequence analysis studies based on Fugu rubripes";  
RL Thesis (2001), University College London, London, United Kingdom.  
DR EMBL; AJ426040; CAD19830.1; -.  
DR GO; GO:0008083; F:growth factor activity; IEA.  
DR InterPro; IPR008996; Cytok IL1\_like.  
DR InterPro; IPR002348; IL1\_HBGF.  
DR Pfam; PF00167; FGF; 1.  
DR PRINTS; PR00262; IL1HBGF.  
DR ProDom; PD000831; IL1\_HBGF; 1.

DR SMART; SM00442; FGF; 1.  
DR PROSITE; PS00247; HBGF\_FGF; 1.  
SQ SEQUENCE 155 AA; 17113 MW; AEFEL2DBDC78FBBE CRC64;  
Query Match 25.8%; Score 618; DB 13; Length 155;  
Best Local Similarity 77.1%; Pred. No. 1.5e-39;  
Matches 118; Conservative 5; Mismatches 30; Indels 0; Gaps 0;  
QY 278 MAAGSITTLPALPEDGGGAPPGHFKDPKRLKCKNGGFFLRIHPDGRVGVREKSDPHI 337  
DB 1 MATGGITTLPTPEDGGGGPPGSGFKDPKRLKCKNGGFFLRIHSDGAVDGTREKTDPHI 60  
QY 338 KLQQAERGVSIGKVCANRYLAMKEDGRLLASKCVTDECFPERLESNNYTYRSKY 397  
DB 61 KLQQAERGVSIGKVCANRYLAMNDRGLFGMKRATDECHFLERLESNNYTYRSKY 120  
QY 398 TSWYVALKRTGQYKLGSKTGPQOKAILFLPMSA 430  
DB 121 PNMFVGLTRTGNYSKSGTKTGPCQKAILFLPMSA 153  
RESULT 8  
Q7ZKZ5 PRELIMINARY; PRT; 154 AA.  
ID Q7ZKZ5  
AC Q7ZKZ5  
DT 01-JUN-2003 (TremBLrel. 24, Created)  
DT 01-JUN-2003 (TremBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)  
DE Fibroblast growth factor 2.  
GN FGF2.  
OS Brachydanio rerio (Zebrafish) (Danio rerio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC Groth C., Lardelli M.;  
RT "Expression analysis of zebrafish fgf2 during embryogenesis.";  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY269790; AAP32155.1; -.  
DR GO; GO:0008083; F:growth factor activity; IEA.  
DR InterPro; IPR008996; Cytok IL1\_like.  
DR InterPro; IPR002348; IL1\_HBGF.  
DR Pfam; PF00167; FGF; 1.  
DR PRINTS; PR00262; IL1HBGF.  
DR ProDom; PD000831; IL1\_HBGF; 1.  
DR SMART; SM00442; FGF; 1.  
DR PROSITE; PS00247; HBGF\_FGF; 1.  
SQ SEQUENCE 154 AA; 17114 MW; 2B28A02B35E1B39B CRC64;  
Query Match 25.3%; Score 607.5; DB 13; Length 154;  
Best Local Similarity 75.2%; Pred. No. 9.4e-39;  
Matches 115; Conservative 10; Mismatches 27; Indels 1; Gaps 1;  
QY 278 MAAGSITTLPALPEDGGGAPPGHFKDPKRLKCKNGGFFLRIHPDGRVGVREKSDPHI 337  
DB 1 MATGGITTLPAAP-DAENSFPAGSFRDPKRLKCKNGGFFLRIHSDGAVDGTREKSDPHI 59  
QY 338 KLQQAERGVSIGKVCANRYLAMKEDGRLLASKCVTDECFPERLESNNYTYRSKY 397  
DB 60 RLQQAERGVSIGKVCANRYLAMNDRGLFGMKRATDECHFLERLESNNYTYRSKY 119  
QY 398 TSWYVALKRTGQYKLGSKTGPQOKAILFLPMSA 430  
DB 120 PNMYVALKRTGQYKSGSKTSPQOKAILFLPMSA 152  
RESULT 9  
Q9BDX1 PRELIMINARY; PRT; 111 AA.  
ID Q9BDX1  
AC Q9BDX1  
DT 01-JUN-2001 (TremBLrel. 17, Created)



DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Basic fibroblast growth factor (Fragment).  
OS Macaca mulatta (Rhesus macaque).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecinae; Macaca.  
OX NCBI\_TaxID=9544;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Sekhon H.S., Keller J.K., Spindel E.R.;  
RT "Alterations in Collagen and Elastin Gene Expression in Fetal  
RT Pulmonary Vessels in Monkeys Following Prenatal Nicotine Exposure: A  
RT Possible Role of alpha7 Nicotinic Acetylcholine Receptor in Persistent  
RT Pulmonary Hypertension.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RE EMBL; AF251270; AAK37962.1; -.  
DR HSP; P09038; 2FGF.  
DR GO; GO:0008083; F: growth factor activity; IEA.  
DR InterPro; IPR008996; Cytok IL1\_like.  
DR InterPro; IPR002348; IL1\_HBGF.  
DR Pfam; PF00167; FGF; 1.  
DR PRINTS; PR00262; IL1HBGF.  
DR ProDom; PD000831; IL1\_HBGF; 1.  
DR SMART; SM00442; FGF; 1.  
DR PROSITE; PS00247; HBGF\_FGF; 1.  
FT NON\_TER 1  
FT NON\_TER 111  
SQ SEQUENCE 111 AA; 12633 MW; EC0967A5261F5487 CRC64;  
Query Match 24.4%; Score 585; DB 6; Length 111;  
Best Local Similarity 100.0%; Pred. No. 3e-37;  
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 320 IHPDGRVGVREKSDPHIKLQQAERGVSIVKGVCANRYLAMKEDGRLLASKCVTDEC 379  
DB 1 IHPDGRVGVREKSDPHIKLQQAERGVSIVKGVCANRYLAMKEDGRLLASKCVTDEC 60  
QY 380 FFERLESNNYNTYRSKYTSWYVALKRTGYKLGSKTGPQKAILFLPMSA 430  
DB 61 FFERLESNNYNTYRSKYTSWYVALKRTGYKLGSKTGPQKAILFLPMSA 111  
RESULT 10  
Q9N1S7  
ID Q9N1S7 PRELIMINARY; PRT; 108 AA.  
AC Q9N1S7;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Basic fibroblast growth factor (Fragment).  
GN BFGF.  
OS Capreolus capreolus (Roe deer).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervioidea;  
OC Cervidae; Odocoileinae; Capreolus.  
OX NCBI\_TaxID=9858;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA TISSUE=testis;  
RC MEDLINE=20532861; PubMed=11078967;  
RX Wagener A., Blottner S., Goritz F., Fickel J.;  
RT "Detection of growth factors in the testis of roe deer (Capreolus  
RT capreolus).";  
RL Anim. Reprod. Sci. 64:65-75(2000).  
DR ENBL; AF152587; AAF73226.1; -.  
DR HSP; P09038; 4FGF.  
DR GO; GO:0008083; F: growth factor activity; IEA.  
DR InterPro; IPR008996; Cytok IL1\_like.  
DR InterPro; IPR002348; IL1\_HBGF.  
DR Pfam; PF00167; FGF; 1.  
DR PRINTS; PR00262; IL1HBGF.  
DR ProDom; PD000831; IL1\_HBGF; 1.

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Basic fibroblast growth factor (Fragment).  
OS Macaca mulatta (Rhesus macaque).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecinae; Macaca.  
OX NCBI\_TaxID=9544;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Sekhon H.S., Keller J.K., Spindel E.R.;  
RT "Alterations in Collagen and Elastin Gene Expression in Fetal  
RT Pulmonary Vessels in Monkeys Following Prenatal Nicotine Exposure: A  
RT Possible Role of alpha7 Nicotinic Acetylcholine Receptor in Persistent  
RT Pulmonary Hypertension.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RE EMBL; AF251270; AAK37962.1; -.  
DR HSP; P09038; 2FGF.  
DR GO; GO:0008083; F: growth factor activity; IEA.  
DR InterPro; IPR008996; Cytok IL1\_like.  
DR InterPro; IPR002348; IL1\_HBGF.  
DR Pfam; PF00167; FGF; 1.  
DR PRINTS; PR00262; IL1HBGF.  
DR ProDom; PD000831; IL1\_HBGF; 1.  
DR SMART; SM00442; FGF; 1.  
DR PROSITE; PS00247; HBGF\_FGF; 1.  
FT NON\_TER 1  
FT NON\_TER 111  
SQ SEQUENCE 111 AA; 12633 MW; EC0967A5261F5487 CRC64;  
Query Match 24.4%; Score 585; DB 6; Length 111;  
Best Local Similarity 100.0%; Pred. No. 3e-37;  
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 320 IHPDGRVGVREKSDPHIKLQQAERGVSIVKGVCANRYLAMKEDGRLLASKCVTDEC 379  
DB 1 IHPDGRVGVREKSDPHIKLQQAERGVSIVKGVCANRYLAMKEDGRLLASKCVTDEC 60  
QY 380 FFERLESNNYNTYRSKYTSWYVALKRTGYKLGSKTGPQKAILFLPMSA 430  
DB 61 FFERLESNNYNTYRSKYTSWYVALKRTGYKLGSKTGPQKAILFLPMSA 111  
RESULT 10  
Q9N1S7  
ID Q9N1S7 PRELIMINARY; PRT; 108 AA.  
AC Q9N1S7;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Basic fibroblast growth factor (Fragment).  
GN BFGF.  
OS Capreolus capreolus (Roe deer).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervioidea;  
OC Cervidae; Odocoileinae; Capreolus.  
OX NCBI\_TaxID=9858;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA TISSUE=testis;  
RC MEDLINE=20532861; PubMed=11078967;  
RX Wagener A., Blottner S., Goritz F., Fickel J.;  
RT "Detection of growth factors in the testis of roe deer (Capreolus  
RT capreolus).";  
RL Anim. Reprod. Sci. 64:65-75(2000).  
DR ENBL; AF152587; AAF73226.1; -.  
DR HSP; P09038; 4FGF.  
DR GO; GO:0008083; F: growth factor activity; IEA.  
DR InterPro; IPR008996; Cytok IL1\_like.  
DR InterPro; IPR002348; IL1\_HBGF.  
DR Pfam; PF00167; FGF; 1.  
DR PRINTS; PR00262; IL1HBGF.  
DR ProDom; PD000831; IL1\_HBGF; 1.

DR SMART; SM00442; FGF; 1.  
DR PROSITE; PS00247; HBGF\_FGF; 1.  
FT NON\_TER 1  
FT NON\_TER 108  
SQ SEQUENCE 108 AA; 12399 MW; 6BC7B7244214567E CRC64;  
Query Match 23.4%; Score 561; DB 6; Length 108;  
Best Local Similarity 98.1%; Pred. No. 2e-35;  
Matches 106; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 319 RIHPDGRVGVREKSDPHIKLQQAERGVSIVKGVCANRYLAMKEDGRLLASKCVTDEC 378  
DB 1 RIHPDGRVGVREKSDPHIKLQQAERGVSIVKGVCANRYLAMKEDGRLLASKCVTDEC 60  
QY 379 FFERLESNNYNTYRSKYTSWYVALKRTGYKLGSKTGPQKAILFL 426  
DB 61 FFERLESNNYNTYRSKYTSWYVALKRTGYKLGSKTGPQKAILFL 108  
RESULT 11  
Q98TD8  
ID Q98TD8 PRELIMINARY; PRT; 125 AA.  
AC Q98TD8;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Fibroblast growth factor-2 (Fragment).  
GN FGF-2.  
OS Cynops pyrrhogaster (Japanese common newt).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Caudata; Salamandridae; Salamandridae; Cynops.  
OX NCBI\_TaxID=8330;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Mizuno N., Hayashi T., Kondoh H., Okamoto M.;  
RT "Cynops fibroblast growth factor-2.";  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB049625; BAB40835.1; -.  
DR HSP; P09038; 1BFF.  
DR GO; GO:0008083; F: growth factor activity; IEA.  
DR InterPro; IPR008996; Cytok IL1\_like.  
DR InterPro; IPR002348; IL1\_HBGF.  
DR Pfam; PF00167; FGF; 1.  
DR PRINTS; PR00262; IL1HBGF.  
DR ProDom; PD000831; IL1\_HBGF; 1.  
DR SMART; SM00442; FGF; 1.  
DR PROSITE; PS00247; HBGF\_FGF; 1.  
FT NON\_TER 1  
SQ SEQUENCE 125 AA; 14244 MW; 5C27F41DC6E60C13 CRC64;  
Query Match 23.4%; Score 561; DB 13; Length 125;  
Best Local Similarity 86.3%; Pred. No. 2.5e-35;  
Matches 107; Conservative 7; Mismatches 10; Indels 0; Gaps 0;  
QY 309 LYCKNGGFFLRIPHGRVGVREKSDPHIKLQQAERGVSIVKGVCANRYLAMKEDGRLL 368  
DB 2 LYCKNGGFFLRINSGKVGAREKSDSYIKLQQAERGVSIVKGVCANRYLAMKEDGRLL 61  
QY 369 LASKCVTDECFFERLESNNYNTYRSKYTSWYVALKRTGYKLGSKTGPQKAILFLPM 428  
DB 62 MALKWITDECFFERLESNNYNTYRSKYTSWYVALKRTGYKLGSKTGPQKAILFLPM 121  
QY 429 SAAS 432  
DB 122 SAKS 125  
RESULT 12  
Q7TPG9  
ID Q7TPG9 PRELIMINARY; PRT; 105 AA.  
AC Q7TPG9;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)	DT	01-DEC-2001 (TrEMBLrel. 19, Created)
DE	FGF2 (Fragment).	DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
GN	FGF2.	DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
OS	Mus musculus (Mouse).	DE	Fibroblast growth factor 2.
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	GN	FGF2.
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	OS	Mus musculus (Mouse).
OX	NCBI_TaxID=10090;	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
RN	[1]	OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RP	SEQUENCE FROM N.A.	OX	NCBI_TaxID=10090;
RC	STRAIN=129/SVJ;	RN	[1]
RA	Foletti A., Vuadens F., Beermann F.;	RP	SEQUENCE FROM N.A.
RT	"Intracellular localization of mouse FGF2.";	RC	STRAIN=FVB/N;
RL	Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.	RA	Dirks R.P., Griep A.E.;
DR	EMBL; AY324451; AAP92385.1; -	RT	"Multiple novel variants of fibroblast growth factor 2 transcripts are expressed in mouse embryos.";
DR	EMBL; AY324449; AAP92385.1; JOINED.	RL	Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AY324450; AAP92385.1; JOINED.	DR	EMBL; AY027558; AAK52310.1; -
FT	NON_TER 1	DR	GO; GO:0008083; F: growth factor activity; IEA.
SQ	SEQUENCE 105 AA; 11945 MW; C24E39323A79D469 CRC64;	DR	InterPro; IPR008996; Cytok IL1-like.
		DR	InterPro; IPR002348; IL1_HBGF.
		DR	Pfam; PF00167; FGF; 1.
		DR	PRINTS; PR00262; IL1HBGF.
		DR	ProDom; PD000831; IL1_HBGF; 1.
		DR	SMART; SM00442; FGF; 1.
		DR	PROSITE; PS00247; HBGF_FGF; 1.
		SQ	SEQUENCE 109 AA; 12388 MW; 61074ADE3303C860 CRC64;
		Query Match	20.2%; Score 484; DB 11; Length 109;
		Best Local Similarity	96.9%; Pred. No. 1.6e-29;
		Matches 93; Conservative	2; Mismatches 1; Indels 0; Gaps 0;
QY	328 GVREKSDPHIKLQQAERGVSIGVCANRYLAMKEDGRLLASKCVTDECFERLESN 387	QY	337 IKLQQAERGVSIGVCANRYLAMKEDGRLLASKCVTDECFERLESNNYTSRK 396
Db	1 GVREKSDPHVQLQQAERGVSIGVCANRYLAMKEDGRLLASKCVTECFERLESN 60	Db	14 IKLQQAERGVSIGVCANRYLAMKEDGRLLASKCVTECFERLESNNYTSRK 73
QY	388 NNTYRSRKYTSWYVALKRTGQYKLGSKTGPQKAILFLPMSAAS 432	QY	397 YTSWYVALKRTGQYKLGSKTGPQKAILFLPMSAAS 432
Db	61 NNTYRSRKYTSWYVALKRTGQYKLGSKTGPQKAILFLPMSAKS 105	Db	74 YSSWYVALKRTGQYKLGSKTGPQKAILFLPMSAKS 109
RESULT 13		RESULT 15	
Q7YRNS		Q25A2	
ID	Q7YRNS PRELIMINARY; PRT; 96 AA.	ID	Q25A2 PRELIMINARY; PRT; 112 AA.
AC	Q7YRNS;	AC	Q25A2;
DT	01-OCT-2003 (TrEMBLrel. 25, Created)	DT	01-DEC-2001 (TrEMBLrel. 19, Created)
DT	01-OCT-2003 (TrEMBLrel. 25, Last sequence update)	DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)	DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	Fibroblast growth factor 2 (Fragment).	DE	Fibroblast growth factor 2.
GN	FGF2.	GN	FGF2.
OS	Sus scrofa (Pig).	OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.	OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxID=9823;	OX	NCBI_TaxID=10090;
RN	[1]	RN	[1]
RP	SEQUENCE FROM N.A.	RP	SEQUENCE FROM N.A.
RC	TISSUE=Endometrium;	RC	STRAIN=FVB/N;
RA	Binspanier R., Welter H.;	RA	Dirks R.P., Griep A.E.;
RT	"Growth factors during porcine endometrial development.";	RT	"Multiple novel variants of fibroblast growth factor 2 transcripts are expressed in mouse embryos.";
RL	Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.	RL	Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AJ577089; CAEL1791.1; -	DR	EMBL; AY027557; AAK52309.1; -
FT	NON_TER 1	DR	GO; GO:0008083; F: growth factor activity; IEA.
FT	NON_TER 96	DR	InterPro; IPR008996; Cytok IL1-like.
SQ	SEQUENCE 96 AA; 11094 MW; FBCA875BED095FA3 CRC64;	DR	InterPro; IPR002348; IL1_HBGF.
		DR	Pfam; PF00167; FGF; 1.
		DR	PRINTS; PR00262; IL1HBGF.
		DR	ProDom; PD000831; IL1_HBGF; 1.
		DR	SMART; SM00442; FGF; 1.
		DR	PROSITE; PS00247; HBGF_FGF; 1.
		SQ	SEQUENCE 112 AA; 12725 MW; B00557ABE0257CCB CRC64;
		Query Match	20.0%; Score 480; DB 11; Length 112;
		Best Local Similarity	96.8%; Pred. No. 3.3e-29;
		Matches 95; Conservative	1; Mismatches 0; Indels 0; Gaps 0;
QY	313 NGGFFLRHPDGRVDGVREKSDPHIKLQQAERGVSIGVCANRYLAMKEDGRLLASK 372	QY	313 NGGFFLRHPDGRVDGVREKSDPHIKLQQAERGVSIGVCANRYLAMKEDGRLLASK 372
Db	1 NGGFFLRHPDGRVDGVREKSDPHIKLQQAERGVSIGVCANRYLAMKEDGRLLASK 60	Db	1 NGGFFLRHPDGRVDGVREKSDPHIKLQQAERGVSIGVCANRYLAMKEDGRLLASK 60
QY	373 CVTDECFERLESNNYTSRKYTSWYVALKRTG 408	QY	373 CVTDECFERLESNNYTSRKYTSWYVALKRTG 408
Db	61 CVTDECFERLESNNYTSRKYTSWYVALKRTG 96	Db	61 CVTDECFERLESNNYTSRKYTSWYVALKRTG 96
RESULT 14		RESULT 14	
Q25A1		Q25A1	
ID	Q25A1 PRELIMINARY; PRT; 109 AA.	ID	Q25A1 PRELIMINARY; PRT; 109 AA.
AC	Q25A1;	AC	Q25A1;

	Matches	92;	Conservative	2;	Mismatches	1;	Indels	0;	Gaps	0;
Qy	338	KLQLQAEERGVS	IKGVCANRYLAMKEDGRLLASKCVTDECFFFERLESNNYNTYRSKY	397						
Db	18	KLQLQAEERGVS	IKGVCANRYLAMKEDGRLLASKCVTDECFFFERLESNNYNTYRSKY	77						
Qy	398	TSMYVALKRTGQYK	LGSKTGPQKAILFLPMSAKS	432						
Db	78	SSWYVALKRTGQYK	LGSKTGPQKAILFLPMSAKS	112						

Search completed: May 3, 2004, 13:11:07  
Job time : 53.1726 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 13:05:13 ; Search time 29.5054 Seconds  
(without alignments)  
1781.157 Million cell updates/sec

Title: US-09-775-964-6  
Perfect score: 1079  
Sequence: 1 GIRLKGKTKGKEDGPGF.....PPGPPKGLGPHGQGRGET 186

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

1: Geneseq1980s:\*

2: Geneseq1990s:\*

3: Geneseq2000s:\*

4: Geneseq2001s:\*

5: Geneseq2002s:\*

6: Geneseq2003as:\*

7: Geneseq2003bs:\*

8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1079	100.0	186	2 AAR34587	Aar34587 Fibronect
2	1079	100.0	186	2 AAW33341	Aaw33341 Oligopept
3	1079	100.0	186	2 AAY05462	Aay05462 Fibronect
4	1079	100.0	464	2 AAR34589	Aar34589 Fibronect
5	1079	100.0	464	2 AAW33342	Aaw33342 Protein u
6	1079	100.0	464	2 AAY05461	Aay05461 Fibronect
7	1079	100.0	1838	2 AAR33257	Aar33257 Human col
8	1079	100.0	1838	7 ADES5566	Ades5566 Human Pro
9	1079	100.0	1838	7 ADES5570	Ades5570 Human Pro
10	1069	99.1	489	2 AAW33343	Aaw33343 Protein u
11	930	86.2	1806	5 AAB84266	Aab84266 Human end
12	930	86.2	1806	5 AABJ05596	Aabj05596 Breast ca
13	930	86.2	1806	6 ABR58545	AbR58545 Human can
14	930	86.2	1806	6 AAB56581	Abu56581 Lung canc
15	835	77.4	454	2 AAR25155	Aar25155 Fibronect
16	834	77.3	167	2 AAR22273	Aar22273 Human col
17	741	68.7	1053	4 AAB85863	Aab85863 Murine ad
18	741	68.7	1739	4 AAG77792	Aag77792 Murine pr
19	737	68.3	1745	4 AAG77793	Aag77793 Human pro
20	737	68.3	1745	5 AAB97234	AbB97234 Novel hum
21	509.5	47.2	1466	4 AAE02534	Aae02534 Bovine al
22	509.5	47.2	1466	4 AAE02533	Aae02533 Bovine al
23	503.5	46.7	357	2 AAR95115	Aar95115 Interveni
24	503.5	46.7	357	2 AAW57646	Aaw57646 Collagen
25	503.5	46.7	357	4 AAB64008	Aab64008 CLP-CB pr

26 503.5 46.7 357 4 AAB72738 Repetitiv  
27 501.5 46.5 772 2 AAR23873 Human alp  
28 501.5 46.5 772 2 AAW09643 Human typ  
29 501.5 46.5 1685 4 ABG04839 Novel hum  
30 501.5 46.5 1693 4 ABG15619 Novel hum  
31 501 46.4 1669 5 Abb57334 Mouse isc  
32 495.5 45.9 1064 2 AAR93254 Collagen-  
33 495.5 45.9 1064 2 AAW57652 Collagen-  
34 495.5 45.9 1065 2 AAR37741 Collagen-  
35 494 45.8 1606 6 ABR40002 Human col  
36 492.5 45.6 1466 4 AAE02537 Porcine a  
37 492 45.6 654 4 AAG63332 Anino aci  
38 491 45.5 1496 5 ABR47419 Breast ca  
39 491 45.5 1496 7 ADB70382 Procollag  
40 491 45.5 1496 7 ADE61180 Human Pro  
41 491 45.5 1496 7 ADE61180 Human Pro  
42 488 45.2 330 2 AAW57645 Collagen-  
43 488 45.2 408 2 AAW07539 Collagen  
44 488 45.2 408 4 AAB64007 CLP prote  
45 488 45.2 408 4 AAB72737 Repetitiv

#### ALIGNMENTS

RESULT 1  
AAR34587  
ID AAR34587 standard; protein; 186 AA.

XX AC AAR34587;

XX XX

DT 14-SEP-1993 (first entry)

XX XX

DE DE

XX XX

KW Polymerase chain reaction; amplify; human; primer; h-ColV; pTYCOLV;

KW intracellular adhesion; PCR; domain; fibronectin; insulin binding; ss.

XX OS Synthetic.

XX XX

PN JP05097698-A.

XX XX

PD 20-APR-1993.

XX PF 14-OCT-1991; 91JP-00291958.

XX PR 14-OCT-1991; 91JP-00291958.

XX XX

PA (TAKI ) TAKARA SHUZO CO LTD.

XX WPI; 1993-164369/20.

XX PT Artificial functional polypeptide having insulin-combining activity -

XX PT includes inter-cellular adhesive domain polypeptide of human fibronectin

XX PT and polypeptide contg. specific aminoacid sequence.

XX PS Disclosure; Page 7-8; 12pp; Japanese.

XX CC The sequences given in AAR34585-90 are polypeptides which were used in

XX CC the production of recombinant polypeptides which contained the

XX CC intracellular adhesive domain of human fibronectin, a spacer and an

XX CC insulin binding domain. These polypeptides have insulin-binding and

XX CC intercellular adhesion activity

XX SQ Sequence 186 AA;

Query Match 100.0%; Score 1079; DB 2; Length 186;

Best Local Similarity 100.0%; Pred. No. 3.7e-69;

Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GIRLKGKTKGKEDGPGFPGKDMGKIDRGIGPPRGEDGPEGPKGGNGGPGPL 60

DB 1 GIRLKGKTKGKEDGPGFPGKDMGKIDRGIGPPRGEDGPEGPKGGNGGPGPL 60

```
QY 61 GPPGKGLGVPLPGYPRQPKSGISGPPGPPGANGKGGCTGKPGPRGQRGPTGPR 120
Db 61 GPPGKGLGVPLPGYPRQPKSGISGPPGPPGANGKGGCTGKPGPRGQRGPTGPR 120

QY 121 GERGPRGITGKPGKNSGGDPAGPPGPRGPNPGQPTGFPKGPPGPKDGLPGHP 180
Db 121 GERGPRGITGKPGKNSGGDPAGPPGPRGPNPGQPTGFPKGPPGPKDGLPGHP 180

QY 181 GORGET 186
Db 181 GORGET 186

RESULT 2
AAW33341
ID AAW33341 standard; protein; 186 AA.
XX AC AAW33341;
XX DT 23-FEB-1998 (first entry)
XX DE Oligopeptide ColV.
XX KW Oligopeptide ColV; target cell; transfection; retroviral vector;
XX KW gene therapy; cancer; viral disease; acquired immunodeficiency syndrome;
XX KW AIDS.
XX OS Synthetic.
XX PN WO9718318-A1.
XX PD 22-MAY-1997.
XX PF 07-NOV-1996; 96WO-JP003254.
XX PR 13-NOV-1995; 95JP-00294382.
XX PR 08-MAR-1996; 96JP-00051847.
XX PA (TAKI ) TAKARA SHUZO CO LTD.
XX PI Asada K, Uemori T, Ueno T, Koyama N, Hashino K, Kato I;
XX WPI; 1997-289294/26.
XX Method for increasing efficacy of gene transfer to target cell using
PT retrovirus - by infection of the target cell in the presence of a
PT substance which binds to the virus and a substance which binds to the
PT target cell.
XX Claim 44; Page 100-101; 194pp; Japanese.
XX The present sequence is the oligopeptide ColV, which was used in the
CC development of a novel method for increasing the efficiency of gene
CC introduction into a target cell using a retroviral vector. The method
CC comprises carrying out viral infection of the target cell in the presence
CC of a retrovirus and target cell binding substance or substances. The
CC method can be used to effectively introduce genes into target cells for
CC the gene therapy of cancer and viral diseases, e.g. AIDS
XX Sequence 186 AA;
Query Match 100.0%; Score 1079; DB 2; Length 186;
Best Local Similarity 100.0%; Pred. No. 3.7e-69;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GIRLKGTKGKGGEDGPPGFKGDMGDKDRGEIGPPGPRGEDGPEGPKRGPNGDGPGPL 60
Db 1 GIRLKGTKGKGGEDGPPGFKGDMGDKDRGEIGPPGPRGEDGPEGPKRGPNGDGPGPL 60

QY 61 GPPGKGLGVPLPGYPRQPKSGISGPPGPPGANGKGGCTGKPGPRGQRGPTGPR 120
Db 61 GPPGKGLGVPLPGYPRQPKSGISGPPGPPGANGKGGCTGKPGPRGQRGPTGPR 120

Query Match 100.0%; Score 1079; DB 2; Length 186;
Best Local Similarity 100.0%; Pred. No. 3.7e-69;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GIRLKGTKGKGGEDGPPGFKGDMGDKDRGEIGPPGPRGEDGPEGPKRGPNGDGPGPL 60
Db 1 GIRLKGTKGKGGEDGPPGFKGDMGDKDRGEIGPPGPRGEDGPEGPKRGPNGDGPGPL 60

QY 61 GPPGKGLGVPLPGYPRQPKSGISGPPGPPGANGKGGCTGKPGPRGQRGPTGPR 120
Db 61 GPPGKGLGVPLPGYPRQPKSGISGPPGPPGANGKGGCTGKPGPRGQRGPTGPR 120
```

```
QY 121 GERGPRGITGKPGKNSGGDPAGPPGPRGPNPGQPTGFPKGPPGPKDGLPGHP 180
Db 121 GERGPRGITGKPGKNSGGDPAGPPGPRGPNPGQPTGFPKGPPGPKDGLPGHP 180

QY 181 GORGET 186
Db 181 GORGET 186

RESULT 3
AAY05462
ID AAY05462 standard; protein; 186 AA.
XX AC AAY05462;
XX DT 07-JUL-1999 (first entry)
XX DE Fibronectin receptor targetting HIV strain ColV.
XX KW Fibronectin receptor; HIV; infection; therapy.
XX OS Unidentified.
XX PN JP10029952-A.
XX PD 03-FEB-1998.
XX PF 16-JUL-1996; 96JP-00185893.
XX PR 16-JUL-1996; 96JP-00185893.
XX PA (TAKI ) TAKARA SHUZO CO LTD.
XX DR WPI; 1998-163674/15.
XX PT Control of human immunodeficiency virus infection - using composition
XX PT comprising replication defective HIV vector.
XX Example 5; Page 23-24; 24pp; Japanese.
XX This sequence represents a fibronectin receptor that can be used in the
CC method of the invention. The method is for the control of human
CC immunodeficiency virus (HIV) infection using a composition which
CC comprises a functional substance which participates in the infection of
CC HIV. The method is used to control HIV-infection
XX Sequence 186 AA;
Query Match 100.0%; Score 1079; DB 2; Length 186;
Best Local Similarity 100.0%; Pred. No. 3.7e-69;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GIRLKGTKGKGGEDGPPGFKGDMGDKDRGEIGPPGPRGEDGPEGPKRGPNGDGPGPL 60
Db 1 GIRLKGTKGKGGEDGPPGFKGDMGDKDRGEIGPPGPRGEDGPEGPKRGPNGDGPGPL 60

QY 61 GPPGKGLGVPLPGYPRQPKSGISGPPGPPGANGKGGCTGKPGPRGQRGPTGPR 120
Db 61 GPPGKGLGVPLPGYPRQPKSGISGPPGPPGANGKGGCTGKPGPRGQRGPTGPR 120

Query Match 100.0%; Score 1079; DB 2; Length 186;
Best Local Similarity 100.0%; Pred. No. 3.7e-69;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 121 GERGPRGITGKPGKNSGGDPAGPPGPRGPNPGQPTGFPKGPPGPKDGLPGHP 180
Db 121 GERGPRGITGKPGKNSGGDPAGPPGPRGPNPGQPTGFPKGPPGPKDGLPGHP 180

QY 181 GORGET 186
Db 181 GORGET 186

RESULT 4
AAR34589
ID AAR34589 standard; protein; 464 AA.
```

XX AAR34589;  
XX 14-SEP-1993 (first entry)  
XX Fibronection domain #3.  
XX Polymerase chain reaction; amplify; human; primer; h-ColV; pTVColV;  
KW intracellular adhesion; PCR; domain; fibronection; insulin binding; ss.  
XX Synthetic.  
XX JF05097698-A.  
XX 20-APR-1993.  
XX 14-OCT-1991; 91JP-00291958.  
XX 14-OCT-1991; 91JP-00291958.  
XX (TAKI ) TAKARA SHUZO CO LTD.  
XX PA  
XX WPI; 1993-164369/20.  
XX  
XX Artificial functional polypeptide having insulin-combining activity -  
PT includes inter-cellular adhesive domain polypeptide of human fibronection  
PT and polypeptide contg. specific aminoacid sequence.  
XX  
XX Disclosure; Page 9-10; 12pp; Japanese.  
XX  
XX The sequences given in AAR34585-90 are polypeptides which were used in  
CC the production of recombinant polypeptides which contained the  
CC intracellular adhesive domain of human fibronection, a spacer and an  
CC insulin binding domain. These polypeptides have insulin-binding and  
CC intercellular adhesion activity  
XX  
SQ Sequence 464 AA;  
Query Match 100.0%; Score 1079; DB 2; Length 464;  
Best Local Similarity 100.0%; Pred. No. 7.9e-69;  
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GIRGLKGTGKEGEDGPPGFKGDMGKGDREIGTPGPRGEDGPEGPKGRGGNGDPGPL 60  
Db 279 GIRGLKGTGKEGEDGPPGFKGDMGKGDREIGTPGPRGEDGPEGPKGRGGNGDPGPL 338  
QY 61 GPGEKGLGVPLPGYPRGQPKSGTGFPGFCANGKGGRTGPKGPRGORGPTGPR 120  
Db 339 GPGEKGLGVPLPGYPRGQPKSGTGFPGFCANGKGGRTGPKGPRGORGPTGPR 398  
QY 121 GERGPRGITGKPGKNSGSDGDPAGPPGERGPNPGQPTGFPKGPDPGKDLPGHP 180  
Db 399 GERGPRGITGKPGKNSGSDGDPAGPPGERGPNPGQPTGFPKGPDPGKDLPGHP 458  
QY 181 QORGET 186  
Db 459 QORGET 464  
RESULT 5  
AAW33342  
ID AAW33342 standard; protein; 464 AA.  
XX  
XX AAW33342;  
XX  
XX 23-FEB-1998 (first entry)  
XX Protein used in development of gene transfer method.  
XX  
XX Target cell; transfection; retroviral vector; gene therapy; cancer;  
KW viral disease; acquired immunodeficiency syndrome; AIDS.  
XX  
XX Synthetic.

XX WO9718318-A1.  
XX 22-MAY-1997.  
XX 07-NOV-1996; 96WO-JP003254.  
XX 13-NOV-1995; 95JP-00294382.  
PR 08-MAR-1996; 96JP-00051847.  
XX  
XX (TAKI ) TAKARA SHUZO CO LTD.  
XX  
XX Asada K, Uemori T, Ueno T, Koyama N, Hashino K, Kato I;  
PI WPI; 1997-289294/26.  
XX  
XX Method for increasing efficacy of gene transfer to target cell using  
PT retrovirus - by infection of the target cell in the presence of a  
PT substance which binds to the virus and a substance which binds to the  
PT target cell.  
XX  
XX Claim 45; Page 100-104; 194pp; Japanese.  
XX  
XX The present sequence was used in the development of a novel method for  
CC increasing the efficiency of gene introduction into a target cell using a  
CC retroviral vector. The method comprises carrying out viral infection of  
CC the target cell in the presence of a retrovirus and target cell binding  
CC substance or substances. The method can be used to effectively introduce  
CC genes into target cells for the gene therapy of cancer and viral  
CC diseases, e.g. AIDS  
XX  
SQ Sequence 464 AA;  
Query Match 100.0%; Score 1079; DB 2; Length 464;  
Best Local Similarity 100.0%; Pred. No. 7.9e-69;  
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GIRGLKGTGKEGEDGPPGFKGDMGKGDREIGTPGPRGEDGPEGPKGRGGNGDPGPL 60  
Db 279 GIRGLKGTGKEGEDGPPGFKGDMGKGDREIGTPGPRGEDGPEGPKGRGGNGDPGPL 338  
QY 61 GPGEKGLGVPLPGYPRGQPKSGTGFPGFCANGKGGRTGPKGPRGORGPTGPR 120  
Db 339 GPGEKGLGVPLPGYPRGQPKSGTGFPGFCANGKGGRTGPKGPRGORGPTGPR 398  
QY 121 GERGPRGITGKPGKNSGSDGDPAGPPGERGPNPGQPTGFPKGPDPGKDLPGHP 180  
Db 399 GERGPRGITGKPGKNSGSDGDPAGPPGERGPNPGQPTGFPKGPDPGKDLPGHP 458  
QY 181 QORGET 186  
Db 459 QORGET 464  
RESULT 6  
AAY05461  
ID AAY05461 standard; protein; 464 AA.  
XX  
XX AAY05461;  
XX  
XX 07-JUL-1999 (first entry)  
XX Fibronection receptor targeting HIV strain C277-Colv.  
XX Fibronection receptor; HIV; infection; therapy.  
XX Unidentified.  
XX JP10029952-A.  
XX  
XX 03-FEB-1998.  
XX  
XX 16-JUL-1996; 96JP-00185893.

```
XX 16-JUL-1996; 96JP-00185893.
XX (TAKI ) TAKARA SHUZO CO LTD.
XX WPI; 1998-163674/15.
XX Control of human immunodeficiency virus infection - using composition
XX comprising replication defective HIV vector.
XX Example 4; Page 22-23; 24pp; Japanese.
XX This sequence represents a fibronectin receptor that can be used in the
XX method of the invention. The method is for the control of human
XX immunodeficiency virus (HIV) infection using a composition which
XX comprises a functional substance which participates in the infection of
XX HIV. The method is used to control HIV-infection
XX Sequence 464 AA;
Query Match 100.0%; Score 1079; DB 2; Length 464;
Best Local Similarity 100.0%; Pred. No. 7.9e-69;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GIRGLKGTGKXGEGDGFPGFKGDMGIKDRGEIGPPRGEDGPGPKRGPGNDPGPL 60
Db 279 GIRGLKGTGKXGEGDGFPGFKGDMGIKDRGEIGPPRGEDGPGPKRGPGNDPGPL 338
QY 61 GPPGEKGLGVPLPGYPRQPKSGIGFPFPGANGKGGRTGPKPGPRQGRGTGPR 120
Db 339 GPPGEKGLGVPLPGYPRQPKSGIGFPFPGANGKGGRTGPKPGPRQGRGTGPR 398
QY 121 GERGRGITGKPGKNSGGDPAGPPGERGNGPQGTGFPKGPDPGKGLPGHP 180
Db 399 GERGRGITGKPGKNSGGDPAGPPGERGNGPQGTGFPKGPDPGKGLPGHP 458
QY 181 GORGET 186
Db 459 GORGET 464
RESULT 7
AAR53257
ID AAR53257 standard; protein; 1838 AA.
XX AAR53257;
XX 12-JAN-1995 (first entry)
XX Human collagen (Type V).
XX Human collagen; alpha 1; V type collagen; placental mRNA.
XX Homo sapiens.
XX Key Location/Qualifiers
XX Peptide 1..38
XX /label= signal_peptide
XX Region 444..538
XX /note= "contains (Gly-X-Y) repeats"
XX Binding-site 645..647
XX /label= RGD
XX /note= "cell adhesion motif"
XX Binding-site 663..665
XX /label= RGD
XX /note= "cell adhesion motif"
XX Domain 897..933
XX /label= heparin_binding_domain
XX Region 1573..1838
XX /label= C-terminal_region
XX /note= "contains 8 Cys residues"
XX JP06105687-A.
```

```
XX 19-APR-1994.
XX 27-DEC-1991; 91JP-00358300.
XX 27-DEC-1991; 91JP-00358300.
XX (TAKI ) TAKARA SHUZO CO LTD.
XX WPI; 1994-163129/20.
XX N-PSDB; AAQ64556.
XX Human collagen V-type gene - is used for diagnosis of human collagen V-
XX type related diseases.
XX Claim 1; Page 6-14; 19pp; Japanese.
XX This amino acid sequence of type V collagen contains several distinct
XX domains including a region comprising repeated (Gly-X-Y) motifs and a
XX central domain containing two copies of the Arg-Gly-Asp cell adhesion
XX motif. The cDNA sequence encoding type V collagen was isolated from a
XX human placental library and will be useful for diagnosis of diseases
XX related to type V collagen
XX Sequence 1838 AA;
Query Match 100.0%; Score 1079; DB 2; Length 1838;
Best Local Similarity 100.0%; Pred. No. 2.6e-68;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GIRGLKGTGKXGEGDGFPGFKGDMGIKDRGEIGPPRGEDGPGPKRGPGNDPGPL 60
Db 799 GIRGLKGTGKXGEGDGFPGFKGDMGIKDRGEIGPPRGEDGPGPKRGPGNDPGPL 858
QY 61 GPPGEKGLGVPLPGYPRQPKSGIGFPFPGANGKGGRTGPKPGPRQGRGTGPR 120
Db 859 GPPGEKGLGVPLPGYPRQPKSGIGFPFPGANGKGGRTGPKPGPRQGRGTGPR 918
QY 121 GERGRGITGKPGKNSGGDPAGPPGERGNGPQGTGFPKGPDPGKGLPGHP 180
Db 919 GERGRGITGKPGKNSGGDPAGPPGERGNGPQGTGFPKGPDPGKGLPGHP 978
QY 181 GORGET 186
Db 979 GORGET 984
RESULT 8
ADE55566
ID ADE55566 standard; protein; 1838 AA.
XX ADE55566;
XX 29-JAN-2004 (first entry)
XX Human Protein P20908, SEQ ID NO 1383.
XX Human; pain; neuronal tissue; gene therapy;
XX spinal segmental nerve injury; chronic constriction injury; CCI;
XX spared nerve injury; SNI; Chung.
XX Homo sapiens.
XX WO2003016475-A2.
XX 27-FEB-2003.
XX 14-AUG-2002; 2002WO-US025765.
XX 14-AUG-2001; 2001US-0312147P.
XX 01-NOV-2001; 2001US-0346382P.
XX 26-NOV-2001; 2001US-0333347P.
XX
```





Db 919 GERGPRGTTGKPGKNSGGDGAGPPGERGNGPQGTGFPKPGPPGKOGLPGHP 978  
181 GORGET 186  
979 GORGET 984  
RESULT 10  
AAW33343  
ID AAW33343 standard; protein; 489 AA.  
AC AAW33343;  
XX  
XX  
DT 23-FEB-1998 (first entry)  
XX  
XX Protein used in development of gene transfer method.  
XX  
XX Target cell; transfection; retroviral vector; gene therapy; cancer;  
KW viral disease; acquired immunodeficiency syndrome; AIDS.  
XX  
XX Synthetic.  
XX  
XX WO9718318-Al.  
XX  
XX 22-MAY-1997.  
XX  
XX 07-NOV-1996; 96WO-JP003254.  
XX  
XX 13-NOV-1995; 95JP-00294382.  
PR 08-MAR-1996; 96JP-00051847.  
XX  
XX (TAKI ) TAKARA SHUZO CO LTD.  
XX  
XX Asada K, Uemori T, Ueno T, Koyama N, Hashino K, Kato I;  
XX  
XX WPI; 1997-289294/26.  
XX  
XX Method for increasing efficacy of gene transfer to target cell using  
PT retrovirus - by infection of the target cell in the presence of a  
PT substance which binds to the virus and a substance which binds to the  
PT target cell.  
XX  
XX Claim 45; Page 104-107; 194pp; Japanese.  
XX  
XX The present sequence was used in the development of a novel method for  
CC increasing the efficiency of gene introduction into a target cell using a  
CC retroviral vector. The method comprises carrying out viral infection of  
CC the target cell in the presence of a retrovirus and target cell binding  
CC substance or substances. The method can be used to effectively introduce  
CC genes into target cells for the gene therapy of cancer and viral  
CC diseases, e.g. AIDS  
XX  
XX Sequence 489 AA;  
SQ  
Query Match 99.1%; Score 1069; DB 2; Length 489;  
Best Local Similarity 100.0%; Pred. No. 4.2e-68;  
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GIRLKGTKGKEDGFPFGKDMGIKGRGEIGPPRGGEDGPEGPKRGGPNGDGPGPL 60  
279 GIRLKGTKGKEDGFPFGKDMGIKGRGEIGPPRGGEDGPEGPKRGGPNGDGPGPL 338  
QY 61 GPPEGKGLGVPLPGYPRGQPKSGIGFPFGPANGKKGGRGTPKPGPRGQGTGPR 120  
339 GPPEGKGLGVPLPGYPRGQPKSGIGFPFGPANGKKGGRGTPKPGPRGQGTGPR 398  
QY 121 GERGPRGTTGKPGKNSGGDGAGPPGERGNGPQGTGFPKPGPPGKOGLPGHP 180  
399 GERGPRGTTGKPGKNSGGDGAGPPGERGNGPQGTGFPKPGPPGKOGLPGHP 458  
QY 181 GORG 184  
|||

Db 459 GORG 462  
RESULT 11  
AAU84266  
ID AAU84266 standard; protein; 1806 AA.  
XX  
XX AC AAU84266;  
XX  
XX 08-MAY-2002 (first entry)  
XX  
XX Human endometrial cancer related protein, COL11A1.  
XX  
XX Human; endometrial cancer; differential expression; DNA microarray;  
KW protein microarray.  
XX  
XX Homo sapiens.  
XX  
XX WO200209573-A2.  
PN  
XX 07-FEB-2002.  
XX  
XX 31-JUL-2001; 2001WO-US024104.  
XX  
XX 31-JUL-2000; 2000US-0221735P.  
PR  
XX (BGHM ) BRIGHAM & WOMENS HOSPITAL INC.  
XX  
XX Mutter GL;  
XX  
XX WPI; 2002-179967/23.  
DR N-PSDB; ABK35486.  
XX  
XX Diagnosing endometrial cancer comprises determining expression of nucleic  
PT acid molecules or expression products that are differentially expressed  
PT in normal and malignant endometrium.  
XX  
XX Claim 33; Page 142-147; 233pp; English.  
XX  
XX The invention relates to diagnosing endometrial cancer in a subject  
CC suspected of having endometrial cancer comprising determining the  
CC expression of a set of nucleic acid molecules or expression products in  
CC an endometrial sample suspected of being cancerous, where the set of  
CC nucleic acid molecules comprises at least 2 nucleic acid molecules  
CC selected from 50 fully defined sequences as given in the specification.  
CC The nucleic acids are used as an array of at least 2 of the 50 nucleic  
CC acids bound to a solid substrate. Also included is a solid-phase protein  
CC microarray comprising at least 2 antibodies or its antigen binding  
CC fragments, that specifically bind at least 2 different polypeptides from  
CC the 50 fully defined sequences as given in the specification, fixed to a  
CC solid substrate. The methods and arrays are useful for the diagnosis of  
CC endometrial cancer, selecting and monitoring treatment regimes and  
CC identification of lead compounds useful for the treatment of endometrial  
CC cancer. The present sequence is one of 50 proteins differentially  
CC expressed between cancerous and non-cancerous samples  
XX  
XX Sequence 1806 AA;  
SQ  
Query Match 86.2%; Score 930; DB 5; Length 1806;  
Best Local Similarity 86.0%; Pred. No. 8.8e-58;  
Matches 160; Conservative 5; Mismatches 21; Indels 0; Gaps 0;  
QY 1 GIRLKGTKGKEDGFPFGKDMGIKGRGEIGPPRGGEDGPEGPKRGGPNGDGPGPL 60  
769 GVRGLKSGKGEKGDGFPFGKDMGLKGRGEVGOIGPRGXDPGPGPKGRAGTGDGPGPS 828  
QY 61 GPPEGKGLGVPLPGYPRGQPKSGIGFPFGPANGKKGGRGTPKPGPRGQGTGPR 120  
829 GQAGEKGLGVPLPGYPRGQPKSGIGFPFGPANGKKGGRGTPKPGPRGQGTGPR 888  
QY 121 GERGPRGTTGKPGKNSGGDGAGPPGERGNGPQGTGFPKPGPPGKOGLPGHP 180  
889 GSRGARGTPGPKGTSGDGPFGPGERGQPGQGVGFPKPGPPGPRMCGPGHP 948  
Db

Qy	181	GORGET 186	
Db	949	GORGET 954	
RESULT 12			
ABJ05596			
ID	ABJ05596	standard; protein; 1806 AA.	
XX	AC	ABJ05596;	
XX	AC		
XX	XX		
DT	14-NOV-2002	(first entry)	
XX	XX		
DE	XX	Breast cancer-associated protein 61.	
XX	XX		
XX	XX	Breast cancer; breast cancer-associated gene sequence; drug development;	
KW	KW	pharmacogenetics; biosensor development.	
KW	XX		
XX	XX	Unidentified.	
OS	XX		
XX	XX		
PN	WO200259377-A2.		
XX	XX		
PD	01-AUG-2002.		
XX	XX		
PD	24-JAN-2002;	2002WO-US002242.	
XX	XX		
PR	24-JAN-2001;	2001US-0263965P.	
XX	XX		
PR	02-FEB-2001;	2001US-0265928P.	
XX	XX		
PR	09-APR-2001;	2001US-00829472.	
XX	XX		
PR	09-APR-2001;	2001US-0282698P.	
XX	XX		
PR	04-MAY-2001;	2001US-0288590P.	
XX	XX		
PR	29-MAY-2001;	2001US-0294443P.	
XX	XX		
PA	(EOSB-) EOS BIOTECHNOLOGY INC.		
XX	XX		
PI	Mack DH, Gish KC, Afar D;		
XX	XX		
PI	WPI; 2002-583738/62.		
XX	XX		
DR	N-PSDB; ABT07753.		
XX	XX		
PT	Detecting a breast cancer-associated transcript in a patient's cell,		
PT	useful for diagnosing breast cancer, comprises contacting a biological		
PT	sample with a polynucleotide that selectively hybridizes with breast		
PT	cancer nucleic acids.		
XX	XX		
PS	Disclosure; Page 403-404; 414pp; English.		
XX	XX		
CC	The invention comprises a method of detecting a breast cancer-associated		
CC	transcript in a cell from a patient. The method of the invention involves		
CC	contacting a biological sample from the patient with a nucleotide that		
CC	hybridizes to one of the 69 breast cancer-associated gene sequences shown		
CC	in the specification. The method of the invention is useful in the		
CC	diagnosis or prognosis of breast cancer, and for detecting genes that are		
CC	up or down-regulated in breast cancer cells. Genes identified by the		
CC	method of the invention can be used in diagnostic purposes and also as		
CC	targets for screening for therapeutic compounds that modulate breast		
CC	cancer (e.g. hormones or antibodies). Identification of genes that are		
CC	over or under expressed in breast cancer can additionally provide high-		
CC	resolution, high-sensitivity datasets which can be used in the areas of		
CC	diagnostics, therapeutics, drug development, pharmacogenetics, protein		
CC	structure and biosensor development. Amino acid sequences ABJ05536 -		
CC	ABJ05604 represent the proteins encoded by the 69 breast cancer-		
CC	associated genes of the invention		
XX	XX		
SQ	Sequence 1806 AA;		
Query Match 86.2%; Score 930; DB 5; Length 1806;			
Best Local Similarity 86.0%; Pred. No. 8.8e-58;			
Matches 160; Conservative 5; Mismatches 21; Indels 0; Gaps 0;			
Qy	1	GTGGLKGTGKGGEDGFPFGKGMGKDRGIBGPPRGEDGPPGKRGNGDPPPL 60	

**QY**

1 GIRLGLTKGEGEDCGFPFKGDMDGKIDRGIEIGPPGRGEDGGPGRKGSGPDGPPL 60  
| : | | | : | | | | | : | | | | | : | | | | | : | | | | |  
CC diagnosing a pathology, e.g. cancer (e.g. cancer of the bone marrow,  
CC therapeutic targets, in particular, the nucleic acid is useful for  
CC bladder, brain, breast, cervix, colon/rectum kidney, lung, ovary,

CC pancreas, prostate, skin and uterus), wounds, ischaemia, heart diseases,  
CC atherosclerosis and endometriosis. The nucleic acid is also useful in  
CC drug screening, particularly for identifying agents for treating these  
CC pathologies  
XX  
SQ Sequence 1806 AA;  
  
Query Match 86.2%; Score 930; DB 6; Length 1806;  
Best Local Similarity 86.0%; Pred. No. 8.8e-58;  
Matches 160; Conservative 5; Mismatches 21; Indels 0; Gaps 0;  
  
QY 1 GIRGLKGTGKEGEDGFPFGKGMGIKGRGEIGPPRGEDGPEGPKRGGPNDDGGL 60  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
769 GVRGLKSGKEGEDGFPFGKGMGLKGRGEVGIQIPRGXDGPEGPKRAGTGDGPGS 828  
  
QY 61 GPPGEKGLGVPLGVPGRQPKSGIGFPPPGANGKGGRTGCKPKGPRGQRTGPR 120  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
829 GQAGEKGLGVPLGVPGRQPKSGTGFPPGANGKGGRTGCKPKGPRGQRTGPR 888  
  
QY 121 GERGPRGITCKPKGKNSGGDGPPGPRGPNPQGTGFPKGPFPKGDGLPGHP 180  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
889 GSRGARGPTGKPKGTSGDGPFGPRGPNPQGTGFPKGPFPKGDGLPGHP 948  
  
QY 181 GORGET 186  
Db :|||||  
949 GORGET 954  
  
RESULT 14  
ABU56581  
ID ABU56581 standard; protein; 1806 AA.  
AC ABU56581;  
XX  
DT 02-APR-2003 (first entry)  
XX  
DE Lung cancer-associated polypeptide #174.  
XX  
KW Lung cancer-associated polypeptide; cytostatic; emphysema;  
KW antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;  
KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;  
KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;  
KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.  
XX  
OS Unidentified.  
XX  
XX WO200286443-A2.  
XX  
XX 31-OCT-2002.  
XX  
XX 18-APR-2002; 2002WO-US012476.  
XX  
PR 18-APR-2001; 2001US-0284770P.  
PR 10-MAY-2001; 2001US-0290492P.  
PR 09-NOV-2001; 2001US-0339245P.  
PR 13-NOV-2001; 2001US-0350666P.  
PR 29-NOV-2001; 2001US-0334370P.  
PR 12-APR-2002; 2002US-0372246P.  
XX  
XX (BOSB-) BOS BIOTECHNOLOGY INC.  
PA  
XX  
XX Aziz N, Murray R;  
XX  
XX WPI; 2003-093161/08.  
DR  
DR N-PSDB; ABX76310.  
DR  
XX  
PT Detecting a lung cancer-associated transcript in a cell from a patient  
PT for treating lung cancer, by contacting a biological sample from the  
PT patient with a polynucleotide that exhibits increased or decreased  
PT expression in lung cancer.  
XX  
XX  
PS Claim 27; Page 322; 453pp; English.  
XX

CC The invention relates to a method for detecting a lung cancer-associated  
CC transcript in a cell from a patient, comprising contacting a biological  
CC sample from the patient with a polynucleotide that selectively hybridises  
CC to a sequence that is at least 80 % identical to a gene that exhibits  
CC increased or decreased expression in lung cancer samples. Lung cancer-  
CC associated polynucleotides and polypeptides are used for identifying a  
CC compound that modulates a lung cancer-associated cell to treat lung  
CC cancer in a patient and for treating a mammal having lung cancer by  
CC administering a modulatory compound identified. The methods are useful  
CC for treating lung cancer, such as small cell lung cancer, non-small cell  
CC lung cancer or other benign or precancerous lesions, e.g. atelectasis,  
CC emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,  
CC hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and  
CC bronchiectasis. The genes, polynucleotides and polypeptides are useful  
CC for diagnostic purposes and as targets for screening for therapeutic  
CC compounds that modulate lung cancer, such as antibodies. Sequences  
CC ABU56408-ABU56745 represent lung cancer-associated polypeptides of the  
CC invention  
XX  
SQ Sequence 1806 AA;  
  
Query Match 86.2%; Score 930; DB 6; Length 1806;  
Best Local Similarity 86.0%; Pred. No. 8.8e-58;  
Matches 160; Conservative 5; Mismatches 21; Indels 0; Gaps 0;  
  
QY 1 GIRGLKGTGKEGEDGFPFGKGMGIKGRGEIGPPRGEDGPEGPKRGGPNDDGGL 60  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
769 GVRGLKSGKEGEDGFPFGKGMGLKGRGEVGIQIPRGXDGPEGPKRAGTGDGPGS 828  
  
QY 61 GPPGEKGLGVPLGVPGRQPKSGIGFPPPGANGKGGRTGCKPKGPRGQRTGPR 120  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
829 GQAGEKGLGVPLGVPGRQPKSGTGFPPGANGKGGRTGCKPKGPRGQRTGPR 888  
  
QY 121 GERGPRGITCKPKGKNSGGDGPPGPRGPNPQGTGFPKGPFPKGDGLPGHP 180  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
889 GSRGARGPTGKPKGTSGDGPFGPRGPNPQGTGFPKGPFPKGDGLPGHP 948  
  
QY 181 GORGET 186  
Db :|||||  
949 GORGET 954  
  
RESULT 15  
AAR25155  
ID AAR25155 standard; protein; 454 AA.  
XX  
XX AAR25155;  
AC AAR25155;  
XX  
XX 25-MAR-2003 (revised)  
DT 17-AUG-1992 (first entry)  
XX  
XX Fibronectin-spacer-collagen XI insulin affinity site fusion protein.  
DE  
XX FN; insulin delivery; cell adhesion.  
KW  
OS Homo sapiens.  
XX  
XX  
FH Key Location/Qualifiers  
FT Domain 1..277  
FT /note= "pro(1239) to Ser(1515) of fibronectin"  
FT Region 278..279  
FT /label= spacer  
FT Domain 280..446  
FT /note= "Pro(261) to Gly(427) of collagen XI"  
XX  
XX JP04054199-A.  
PN  
XX  
XX 21-FEB-1992.  
PD  
XX 20-JUN-1990; 90JP-00159624.  
PF  
XX 20-JUN-1990; 90JP-00159624.  
PR

```

XX (TAKI ) TAKARA SHUZO CO LTD.
XX
XX WPI; 1992-110872/14.
XX
XX Polypeptide for new insulin delivery system - contg. sequence of human
XX fibronectin cell adhesion domain and sequence of human vitronectin or
XX collagen XI insulin binding site.
XX
XX Disclosure; Page 3 and Fig 2; 13pp; Japanese.
XX
XX This fusion protein is a preferred sequence covered by the invention. It
XX comprises amino acids 1239 to 1515 of human fibronectin linked to amino
XX acids 261 to 427 of human collagen XI via an optional spacer. The C-
XX terminal sequence is derived from lacZ. See AAQ26040-5 and AAR22274.
XX (Updated on 25-MAR-2003 to correct PA field.)
XX
XX Sequence 454 AA;
SQ
Query Match 77.4%; Score 835; DB 2; Length 454;
Best Local Similarity 85.1%; Pred. No. 1.4e-51;
Matches 143; Conservative 4; Mismatches 21; Indels 0; Gaps 0;
QY 18 PGFKGDMGKIGDRGEIGPPGRCGEDGEGPKGRCGNPDGPIGPPGPKGKLGVPGLPGY 77
Db 280 PGFKGDMGLKGRGEVQIGPRGXDGPEGPKGRAGTGDGPGSQAGEKGLGVPGLPGY 339
QY 78 PGKQPKGSGIGFPGFPGANGKGGRTGPKGPRGQRTGPRGPRGRTGKPKGKN 137
Db 340 PGKQPKGSTGTFPGFPGANGKKGARGVAGFPGFPGQRTGPRGSRGARGTGPGRPKGT 399
QY 138 SGGDGPAGPPGPRGNPGQGTGTPGPKGPPGPKDGLFPGHPCQGE 185
Db 400 SGGDGPAGPPGPRGNPGQGTGTPGPKGPPGPKDGLFPGHPCQGE 447

```

Search completed: May 3, 2004, 13:09:46  
Job time : 31.5054 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 13:07:19 ; Search time 9.54588 Seconds  
(without alignments)  
1005.924 Million cell updates/sec

Title: US-09-775-964-6  
Perfect score: 1079  
Sequence: 1 GINGLGTGKEKEDGPGF.....PPGPPGKGLPHGPGQGET 186

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/2/iaa/5A COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1079	100.0	186	4	US-09-366-009-6
2	1079	100.0	186	4	US-08-809-156B-6
3	1079	100.0	464	2	US-08-836-854-19
4	1079	100.0	464	4	US-09-366-009-7
5	1079	100.0	464	4	US-08-809-156B-7
6	1069	99.1	489	4	US-09-366-009-8
7	1069	99.1	489	4	US-08-809-156B-8
8	834	77.3	446	2	US-08-836-854-15
9	503.5	46.7	357	1	US-07-609-716-66
10	503.5	46.7	357	1	US-08-642-255-33
11	503.5	46.7	357	3	US-08-475-411A-66
12	503.5	46.7	357	3	US-08-478-029A-66
13	495.5	45.9	1064	1	US-08-642-255-62
14	488	45.2	330	1	US-08-642-255-32
15	488	45.2	408	1	US-07-609-716-65
16	488	45.2	408	3	US-08-475-411A-65
17	488	45.2	408	3	US-08-478-029A-65
18	486.5	45.1	1057	3	US-08-931-820-4
19	486.5	45.1	1078	3	US-08-963-825-21
20	486.5	45.1	1078	4	US-09-500-811-21
21	486.5	45.1	1078	4	US-09-570-573-21
22	486.5	45.1	1078	4	US-09-548-608-21
23	486	45.0	1024	3	US-08-931-820-2
24	486	45.0	1366	3	US-08-963-825-19
25	486	45.0	1366	4	US-09-500-811-19
26	486	45.0	1366	4	US-09-570-573-19
27	486	45.0	1366	4	US-09-548-608-19

28 483.5 44.8 252 1 US-08-642-255-61 Sequence 61, Appl  
29 483 44.8 492 4 US-08-468-996-12 Sequence 12, Appl  
30 481.5 44.6 532 1 US-08-494-168-9 Sequence 9, Appl  
31 479.5 44.4 471 2 US-08-399-889-24 Sequence 24, Appl  
32 479.5 44.4 471 3 US-09-167-364-24 Sequence 24, Appl  
33 479.5 44.4 471 3 US-09-439-897-2 Sequence 2, Appl  
34 478.5 44.3 1017 4 US-08-468-996-10 Sequence 10, Appl  
35 478.5 44.3 1057 3 US-08-931-820-1 Sequence 1, Appl  
36 478.5 44.3 1341 3 US-08-963-825-18 Sequence 18, Appl  
37 478.5 44.3 1341 4 US-09-500-811-18 Sequence 18, Appl  
38 478.5 44.3 1341 4 US-09-570-573-18 Sequence 18, Appl  
39 478.5 44.3 1341 4 US-09-548-608-18 Sequence 18, Appl  
40 478.5 44.3 1461 4 US-09-585-887-9 Sequence 9, Appl  
41 478.5 44.3 1461 4 US-09-289-578-9 Sequence 9, Appl  
42 478.5 44.3 1366 4 US-09-331-347C-21 Sequence 21, Appl  
43 478 44.3 1366 4 US-09-585-887-10 Sequence 10, Appl  
44 478 44.3 1366 4 US-09-289-578-10 Sequence 10, Appl  
45 477 44.2 504 3 US-09-219-849-3 Sequence 3, Appl

#### ALIGNMENTS

RESULT 1  
US-09-366-009-6  
; Sequence 6, Application US/09366009  
; Patent No. 6426042  
; GENERAL INFORMATION:  
; APPLICANT: Asada, Kiyozo  
; Uemori, Takashi  
; Koyama, No. 6426042uto  
; Hashino, Kimikazu  
; Kato, Ikunoshin  
TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: WEISER & ASSOCIATES  
STREET: 230 South Fifteenth Street, Suite 500  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19102

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/366,009  
FILING DATE: 02-Aug-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/809,156  
FILING DATE: <Unknown>  
APPLICATION NUMBER: JP 294382/1995  
FILING DATE: 13-NOV-1995  
APPLICATION NUMBER: JP 051847/1996  
FILING DATE: 08-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Weiser, Gerard J.  
REGISTRATION NUMBER: 19,763  
REFERENCE/DOCKET NUMBER: 977.6507P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-875-8383  
TELEFAX: 215-875-8394  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 186 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear

```
;
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-366-009-6

Query Match          100.0%; Score 1079; DB 4; Length 186;
Best Local Similarity 100.0%; Pred. No. 1.6e-76;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GIRGLKGTGKGEDEGFGFKGDMGIKDRGEIGPPRGEDGPEGPKRGCGPNCDGPPL 60
   |||||
Db 1 GIRGLKGTGKGEDEGFGFKGDMGIKDRGEIGPPRGEDGPEGPKRGCGPNCDGPPL 60
   |||||
QY 61 GPPGEKGLGVPLGPGYPRQPKSGISGPPFGGANGKGGRTGPKDPPRQGRGPTGPR 120
   |||||
Db 61 GPPGEKGLGVPLGPGYPRQPKSGISGPPFGGANGKGGRTGPKDPPRQGRGPTGPR 120
   |||||
QY 121 GERGPRGITGKPGKNSGGDGPAGPPGERGPNQGGTGFPGKGGPPGKGLPGHP 180
   |||||
Db 121 GERGPRGITGKPGKNSGGDGPAGPPGERGPNQGGTGFPGKGGPPGKGLPGHP 180
   |||||
QY 181 GORGET 186
   |||||
Db 181 GORGET 186
   |||||

RESULT 2
US-08-809-156B-6
; Sequence 6, Application US/08809156B
; Patent No. 6472204
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; APPLICANT: Uemori, Takashi
; APPLICANT: Ueno, Takashi
; APPLICANT: Koyama, No. 6472204uto
; APPLICANT: Hashino, Kimikazu
; APPLICANT: Kato, Ikunoshin
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
; TITLE OF INVENTION: CELLS WITH RETROVIRUS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/809,156B
; FILING DATE: 07-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03254
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 294382/1995
; FILING DATE: 13-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 051847/1996
; FILING DATE: 08-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 977.6507P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 186 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

;
; LENGTH: 186 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-809-156B-6

Query Match          100.0%; Score 1079; DB 4; Length 186;
Best Local Similarity 100.0%; Pred. No. 1.6e-76;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GIRGLKGTGKGEDEGFGFKGDMGIKDRGEIGPPRGEDGPEGPKRGCGPNCDGPPL 60
   |||||
Db 1 GIRGLKGTGKGEDEGFGFKGDMGIKDRGEIGPPRGEDGPEGPKRGCGPNCDGPPL 60
   |||||
QY 61 GPPGEKGLGVPLGPGYPRQPKSGISGPPFGGANGKGGRTGPKDPPRQGRGPTGPR 120
   |||||
Db 61 GPPGEKGLGVPLGPGYPRQPKSGISGPPFGGANGKGGRTGPKDPPRQGRGPTGPR 120
   |||||
QY 121 GERGPRGITGKPGKNSGGDGPAGPPGERGPNQGGTGFPGKGGPPGKGLPGHP 180
   |||||
Db 121 GERGPRGITGKPGKNSGGDGPAGPPGERGPNQGGTGFPGKGGPPGKGLPGHP 180
   |||||
QY 181 GORGET 186
   |||||
Db 181 GORGET 186
   |||||

RESULT 3
US-08-836-854-19
; Sequence 19, Application US/08836854
; Patent No. 5824547
; GENERAL INFORMATION:
; APPLICANT: HASHINO, Kimikazu
; APPLICANT: MATSUSHITA, Hideyuki
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: METHOD OF PRODUCTION OF TRANSFECTED CELLS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W. Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,854
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP95/02425
; FILING DATE: 29-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 317721/1994
; FILING DATE: 29-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: HASHINO=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 464 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```



```

; MOLECULE TYPE: peptide
; US-09-366-009-7
;
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
; US-09-366-009-7
;
; MOLECULE TYPE: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
; US-09-366-009-7
;
Query Match 100.0%; Score 1079; DB 4; Length 464;
Best Local Similarity 100.0%; Pred. No. 3.8e-76;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GIRLKGTTGKGEKEDGFPFGKDMGKIDRGIGIPPPRGEDGPGKGRGPNGDGPGPL 60
DB 279 GIRLKGTTGKGEKEDGFPFGKDMGKIDRGIGIPPPRGEDGPGKGRGPNGDGPGPL 338
QY 61 GPPGEGKGLGVPLGPGYQGRQPKGSIGFPPGPGANGKGGRTGPKGPGRGQGTGPR 120
DB 339 GPPGEGKGLGVPLGPGYQGRQPKGSIGFPPGPGANGKGGRTGPKGPGRGQGTGPR 398
QY 121 GERGPRGITTCKPGKNSGGDGPAGPPGGRGPNQPGTGFPPGKPPGPKDGLPGHP 180
DB 399 GERGPRGITTCKPGKNSGGDGPAGPPGGRGPNQPGTGFPPGKPPGPKDGLPGHP 458
QY 181 GQRGET 186
DB 459 GQRGET 464

RESULT 4
US-09-366-009-7
; Sequence 7, Application US/09366009
; Patent No. 6426042
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; Uemori, Takashi
; Ueno, Takashi
; Koyama, No. 6426042uto
; Hashino, Kimikazu
; Kato, Ikunoshin
;
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
;
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/366,009
; FILING DATE: 02-Aug-1999
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/809,156
; FILING DATE: <Unknown>
; APPLICATION NUMBER: JP 294382/1995
; FILING DATE: 13-NOV-1995
; APPLICATION NUMBER: JP 051847/1996
; FILING DATE: 08-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 977.6507P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
;
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 464 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>

```



REFERENCE/DOCKET NUMBER: 977.6507P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-875-8383  
TELEFAX: 215-875-8394  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 489 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-809-156B-8

Query Match 99.1%; Score 1069; DB 4; Length 489;  
Best Local Similarity 100.0%; Pred. No. 2.4e-75;  
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GIRLKGKTEKEDGPPGFKGDMGKGRGEGTGGPRGEGDPEGPKGGGNGDGPGL 60  
DB 279 GIRLKGKTEKEDGPPGFKGDMGKGRGEGTGGPRGEGDPEGPKGGGNGDGPGL 338  
QY 61 GPPEKKGKLVGVLGYPGRQKSGISGFFGFCANGKGGRTGPKGPRGGRGRTGPR 120  
DB 339 GPPEKKGKLVGVLGYPGRQKSGISGFFGFCANGKGGRTGPKGPRGGRGRTGPR 398  
QY 121 GERGPRGRTGPKGKNSGGDGPAGPGRGPNQPGTGFPGPKGGKDGLPGHP 180  
DB 399 GERGPRGRTGPKGKNSGGDGPAGPGRGPNQPGTGFPGPKGGKDGLPGHP 458  
QY 181 GQKG 184  
DB 459 GQKG 462

RESULT 8  
US-08-836-854-15  
Sequence 15, Application US/08836854  
Patent No. 5824547  
GENERAL INFORMATION:  
APPLICANT: HASHINO, Kimikazu  
APPLICANT: MATSUSHITA, Hideyuki  
APPLICANT: KATO, Ikunoshi  
TITLE OF INVENTION: METHOD OF PRODUCTION OF TRANSFECTED CELLS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street N.W. Ste. 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/836,854  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP95/02425  
FILING DATE: 29-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 317721/1994  
FILING DATE: 29-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Browdy, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: HASHINO-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-5197  
TELEFAX: (202) 737-3528

INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 446 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-836-854-15

Query Match 77.3%; Score 834; DB 2; Length 446;  
Best Local Similarity 85.6%; Pred. No. 2.8e-57;  
Matches 143; Conservative 3; Mismatches 21; Indels 0; Gaps 0;  
QY 18 PGFKGDMGKGRGEGTGGPRGEGDPEGPKGGGNGDGPGLGPPGKKGKLVGVLGPGY 77  
DB 280 PGFKGDMGKGRGEGTGGPRGEGDPEGPKGGGNGDGPGLGPPGKKGKLVGVLGPGY 339  
QY 78 PGROGKSGTGGPRGEGDPEGPKGGGNGDGPGLGPPGKKGKLVGVLGPGY 137  
DB 340 PGROGKSGTGGPRGEGDPEGPKGGGNGDGPGLGPPGKKGKLVGVLGPGY 399  
QY 138 SGGDGPAGPGRGPNQPGTGFPGPKGGKDGLPGHPGQKG 184  
DB 400 SGGDGPAGPGRGPNQPGTGFPGPKGGKDGLPGHPGQKG 446

RESULT 9  
US-07-609-716-66  
Sequence 66, Application US/07609716  
Patent No. 5514581  
GENERAL INFORMATION:  
APPLICANT: Ferrari, Franco A.  
APPLICANT: Cappello, Joseph  
TITLE OF INVENTION: Functional Recombinantly Prepared  
TITLE OF INVENTION: Synthetic Protein Polymer  
NUMBER OF SEQUENCES: 118  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: US  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/609,716  
FILING DATE: 06-NOV-1990  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Rowland, Bertram I.  
REGISTRATION NUMBER: 20015  
REFERENCE/DOCKET NUMBER: A-55186-3/BIR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-781-1989  
TELEFAX: 415-398-3249  
INFORMATION FOR SEQ ID NO: 66:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 357 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-609-716-66

Query Match 46.7%; Score 503.5; DB 1; Length 357;  
Best Local Similarity 49.5%; Pred. No. 7e-32;  
Matches 105; Conservative 4; Mismatches 76; Indels 27; Gaps 3;  
QY 1 GIRLKGKTEKEDGPPGFKGDMGKGRGEGTGGPRGEGDPEGPKGGGNGDGP 57





## RESULT 14

```

US-08-642-255-32
; Sequence 32, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 330 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-642-255-32

```

Query Match	45.2%	Score 488;	DB 1;	Length 330;
Best Local Similarity	52.1%;	Pred.No.1e-30;		
Matches 99;	Conservative 1;	Mismatches 84;	Indels 6;	Gaps 1;
QY	1	GIRGLKTKKEGBDGGPPGFKDGMGIKGDRCEIGPPGPRGEDGEGPKGRGPNCDGPGL	60	
Db	61	GPFGPAGPVGSGPAGPFPFPFGPPGPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP	120	
QY	61	GPPEKEKGLVGPLGYPGRQPKSGSIGFPFGANCEKGRGTGKPGPRGQRGTGPR	120	
Db	121	GPFGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP	180	
QY	121	GERGP-----RGITGKEGPKNGSGGDPAGPPGERGPNQGPTGPGPKGPPGPKGD	174	
Db	181	GPAGPVGSGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP	240	
QY	175	GLFCHPGQQRG	184	
Db	241	GPFGCAPGPPG	250	

RESULT 15

RESUL 13  
US-07-609-716-65  
; Sequence 65, Application US/07609716  
; Patent No. 5514581  
; GENERAL INFORMATION:  
; APPLICANT: Ferrari, Franco A.  
; APPLICANT: Cappello, Joseph

```

; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 118
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fleht, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/609,716
; FILING DATE: 06-NOV-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A-55186-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-609-716-65

Query Match 45.2%; Score 488; DB 1; Length 408;
Best Local Similarity 52.1%; Pred. No. 1.2e-30;
Matches 99; Conservative 1; Mismatches 84; Indels 6; Gaps 1

Qy 1 GIRGLKTKGEKEDGPPGFGKDMIGIKGDRGEIGPPGPRGEDGPEGPKRGCGPNGDG 60
Db 61 GPPGAGPVGSPGAGPPGPPGPPGPGAGCPGPPGPPGPPGAGPVGSPGAGPPGPP 120
Qy 61 GPPGEEKLGVPLGVPGRGKPGKSTGTFGFPDANGKEKGRGTFPGKPGRGKRGPTGR 120
Db 121 GPPGPPGAPGPPGPPGPPGPPGPPGAGPVGSPGAGPPGPPGPPGPPGPPGPP 180
Qy 121 GERGP-----RGITGPKGPKNGSGGDPGAPGPPGERGPNQGTFGPPGKPGPPGPKD 174
Db 181 GPAGPVGSPGAPGPPGPPGPPGPPGPPGAGPPGPPGPPGPPGPPGPPGPPGPP 240
Qy 175 GLFHPGQRG 184
Db 241 GPPGAGPPG 250

Search completed: May 3, 2004, 13:12:35
Job time : 10.5459 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 13:11:14 ; Search time 22.2737 Seconds  
(without alignments)  
2314.713 Million cell updates/sec

Title: US-09-775-964-6  
Perfect score: 1079  
Sequence: 1 GRLGLKGTGKGEDEGPGF.....PPGPPKGLGHPGQGRGET 186

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1138120 seqs, 277189581 residues

Total number of hits satisfying chosen parameters: 1138120

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1079	100.0	186	10 US-09-775-964-6	Sequence 6, Appli
2	1079	100.0	464	10 US-09-775-964-7	Sequence 7, Appli
3	1069	99.1	489	10 US-09-775-964-8	Sequence 8, Appli
4	930	86.2	1806	9 US-09-919-497-56	Sequence 56, Appli
5	930	86.2	1806	12 US-10-058-270A-122	Sequence 122, App
6	741	68.7	1739	10 US-09-795-061-2	Sequence 2, Appli
7	737	68.3	1745	10 US-09-795-061-4	Sequence 4, Appli
8	509.5	47.2	1466	15 US-10-402-089-4	Sequence 4, Appli
9	509.5	47.2	1466	15 US-10-402-089-6	Sequence 6, Appli
10	509.5	47.2	1466	15 US-10-402-072A-4	Sequence 4, Appli
11	509.5	47.2	1466	15 US-10-402-072A-6	Sequence 6, Appli
12	492.5	45.6	1466	15 US-10-402-089-12	Sequence 12, Appli
13	492.5	45.6	1466	15 US-10-402-072A-12	Sequence 12, Appli
14	491	45.5	1496	14 US-10-177-293-70	Sequence 70, Appli
15	491	45.5	1496	14 US-10-301-822-35	Sequence 35, Appli

16	491	45.5	1496	15	US-10-236-031B-74	Sequence 74, Appli
17	487	45.1	1449	15	US-10-402-089-8	Sequence 8, Appli
18	487	45.1	1449	15	US-10-402-072A-8	Sequence 8, Appli
19	486.5	45.1	1078	14	US-10-058-124-21	Sequence 21, Appli
20	486.5	45.1	1466	12	US-09-918-715-226	Sequence 226, App
21	486.5	45.1	1466	12	US-10-257-021-72	Sequence 72, Appli
22	486.5	45.1	1466	14	US-10-177-293-68	Sequence 68, Appli
23	486.5	45.1	1466	14	US-10-301-822-33	Sequence 33, Appli
24	486	45.0	1366	12	US-10-257-021-76	Sequence 76, Appli
25	486	45.0	1366	14	US-10-058-124-19	Sequence 19, Appli
26	486	45.0	1366	15	US-10-402-089-10	Sequence 10, Appli
27	486	45.0	1366	15	US-10-402-072A-10	Sequence 10, Appli
28	486	45.0	1669	12	US-09-918-715-252	Sequence 252, App
29	486	45.0	1669	15	US-10-372-683-8	Sequence 8, Appli
30	484	44.9	1497	14	US-10-157-031-128	Sequence 128, App
31	483.5	44.8	1212	15	US-10-331-496A-26	Sequence 26, Appli
32	480	44.5	1626	14	US-10-202-167-2	Sequence 2, Appli
33	479.5	44.4	638	15	US-10-138-588-42	Sequence 42, Appli
34	479.5	44.4	1516	14	US-10-060-036-166	Sequence 166, App
35	479.5	44.4	1516	15	US-10-431-642-3	Sequence 3, Appli
36	478.5	44.3	510	12	US-10-232-175-26	Sequence 26, Appli
37	478.5	44.3	638	13	US-10-001-887-108	Sequence 108, App
38	478.5	44.3	662	12	US-10-232-175-33	Sequence 33, Appli
39	478.5	44.3	703	14	US-10-219-449-4	Sequence 4, Appli
40	478.5	44.3	714	14	US-10-233-885-44	Sequence 44, Appli
41	478.5	44.3	714	14	US-10-231-581-44	Sequence 44, Appli
42	478.5	44.3	714	15	US-10-326-508A-15	Sequence 15, Appli
43	478.5	44.3	717	14	US-10-219-449-2	Sequence 2, Appli
44	478.5	44.3	733	15	US-10-138-588-40	Sequence 40, Appli
45	478.5	44.3	1341	14	US-10-058-124-18	Sequence 18, Appli

ALIGNMENTS

RESULT 1  
US-09-775-964-6  
; Sequence 6, Application US/09775964  
; Publication No. US20030087437A1  
; GENERAL INFORMATION:  
; APPLICANT: Asada, Kiyozo  
; Uemori, Takashi  
; Ueno, Takashi  
; Koyama, No. US20030087437A1uto  
; Hashino, Kimikazu  
; Kato, Ikunoshin  
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET  
; CELLS WITH RETROVIRUS  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: WEISER & ASSOCIATES  
; STREET: 230 South Fifteenth Street, Suite 500  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19102  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/775,964  
; FILING DATE: 20-Feb-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/366,009  
; FILING DATE: 02-Aug-1999  
; APPLICATION NUMBER: 08/809,156  
; FILING DATE: <Unknown>  
; APPLICATION NUMBER: JP 294382/1995  
; FILING DATE: 13-NOV-1995  
; APPLICATION NUMBER: JP 051847/1996

```
;
; FILING DATE: 08-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 977.6507P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 186 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-775-964-6

Query Match 100.0%; Score 1079; DB 10; Length 186;
Best Local Similarity 100.0%; Pred. No. 2.5e-65;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GIRGLKGTGKGEKGDGFFGFKGDMGKIGDRGIEIGPPRGEDGPGPKRGPGNGDPGL 60
Db 1 GIRGLKGTGKGEKGDGFFGFKGDMGKIGDRGIEIGPPRGEDGPGPKRGPGNGDPGL 60

QY 61 GPPGKGLGVPLGYPGRQPKSIGFFPGGANGKGGGRTGKPGRQGRGTGPR 120
Db 61 GPPGKGLGVPLGYPGRQPKSIGFFPGGANGKGGGRTGKPGRQGRGTGPR 120

QY 121 GERGPRGITGKPGKNSGCDGPPGPRGPNPGQGTGFPKPGKPGPPGKGLPGHP 180
Db 121 GERGPRGITGKPGKNSGCDGPPGPRGPNPGQGTGFPKPGKPGPPGKGLPGHP 180

QY 181 GORGET 186
Db 181 GORGET 186

RESULT 2
US-09-775-964-7
; Sequence 7, Application US/09775964
; Publication No. US20030087437A1
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; Uemori, Takashi
; Koyama, No. US20030087437Aluto
; Hashino, Kimikazu
; Kato, Ikunoshin
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
; CELLS WITH RETROVIRUS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/775,964
; FILING DATE: 20-Feb-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/366,009
; FILING DATE: 02-Aug-1999
; APPLICATION NUMBER: 08/809,156

;
; FILING DATE: <Unknown>
; APPLICATION NUMBER: JP 294382/1995
; FILING DATE: 13-NOV-1995
; APPLICATION NUMBER: JP 051847/1996
; FILING DATE: 08-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 977.6507P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 464 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-775-964-7

Query Match 100.0%; Score 1079; DB 10; Length 464;
Best Local Similarity 100.0%; Pred. No. 5.6e-65;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GIRGLKGTGKGEKGDGFFGFKGDMGKIGDRGIEIGPPRGEDGPGPKRGPGNGDPGL 60
Db 279 GIRGLKGTGKGEKGDGFFGFKGDMGKIGDRGIEIGPPRGEDGPGPKRGPGNGDPGL 338

QY 61 GPPGKGLGVPLGYPGRQPKSIGFFPGGANGKGGGRTGKPGRQGRGTGPR 120
Db 339 GPPGKGLGVPLGYPGRQPKSIGFFPGGANGKGGGRTGKPGRQGRGTGPR 398

QY 121 GERGPRGITGKPGKNSGCDGPPGPRGPNPGQGTGFPKPGKPGPPGKGLPGHP 180
Db 399 GERGPRGITGKPGKNSGCDGPPGPRGPNPGQGTGFPKPGKPGPPGKGLPGHP 458

QY 181 GORGET 186
Db 459 GORGET 464

RESULT 3
US-09-775-964-8
; Sequence 8, Application US/09775964
; Publication No. US20030087437A1
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; Uemori, Takashi
; Ueno, Takashi
; Koyama, No. US20030087437Aluto
; Hashino, Kimikazu
; Kato, Ikunoshin
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
; CELLS WITH RETROVIRUS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/775,964
; FILING DATE: 20-Feb-2001
; CLASSIFICATION: <Unknown>
; APPLICATION NUMBER: US/09/366,009
; FILING DATE: 02-Aug-1999
; APPLICATION NUMBER: 08/809,156
```



PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/366,009  
; FILING DATE: 02-Aug-1999  
; APPLICATION NUMBER: 08/809,156  
; FILING DATE: <Unknown>  
; APPLICATION NUMBER: JP 294382/1995  
; FILING DATE: 13-NOV-1995  
; APPLICATION NUMBER: JP 051847/1996  
; FILING DATE: 08-MAR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weiser, Gerard J.  
; REGISTRATION NUMBER: 19,763  
; REFERENCE/DOCKET NUMBER: 977.6507P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-875-8383  
; TELEFAX: 215-875-8394  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 489 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: <Unknown>  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
US-09-775-964-8

Query Match 99.1%; Score 1069; DB 10; Length 489;  
Best Local Similarity 100.0%; Pred. No. 2.8e-64;  
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GIRGLKGTGKSGDGGPGFKGDMGIKDRGEIGPRGPRGEDGEGPKGSGGNGDGPPL 60  
DB 279 GIRGLKGTGKSGDGGPGFKGDMGIKDRGEIGPRGPRGEDGEGPKGSGGNGDGPPL 338  
QY 61 GPPGEKGLGVPLGPGYGRQPKSGISGFPFGPANGKGGRTGPKGPRGGRGPTGPR 120  
DB 339 GPPGEKGLGVPLGPGYGRQPKSGISGFPFGPANGKGGRTGPKGPRGGRGPTGPR 398  
QY 121 GERPRGRTGPKGSGGDPAGPGRGPNQPGTFFPKGPPGPKGDKGLPGHP 180  
DB 399 GERPRGRTGPKGSGGDPAGPGRGPNQPGTFFPKGPPGPKGDKGLPGHP 458  
QY 181 GORG 184  
DB 459 GORG 462

RESULT 4  
US-09-919-497-56  
; Sequence 56, Application US/09919497  
; Patent No. US20020106662A1  
; GENERAL INFORMATION:  
; APPLICANT: Mutter, George L.  
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER  
; FILE REFERENCE: B0801/7225  
; CURRENT APPLICATION NUMBER: US/09/919,497  
; CURRENT FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: US 60/221,735  
; PRIOR FILING DATE: 2000-07-31  
; NUMBER OF SEQ ID NOS: 100  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 56  
; LENGTH: 1806  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (758)..(758)  
; OTHER INFORMATION: Xaa = any amino acid  
; NAME/KEY: UNSURE  
; LOCATION: (809)..(809)  
; OTHER INFORMATION: Xaa = any amino acid  
US-09-919-497-56

Query Match 86.2%; Score 930; DB 9; Length 1806;  
Best Local Similarity 86.0%; Pred. No. 1.8e-54;  
Matches 160; Conservative 5; Mismatches 21; Indels 0; Gaps 0;  
QY 1 GIRGLKGTGKSGDGGPGFKGDMGIKDRGEIGPRGPRGEDGEGPKGSGGNGDGPPL 60  
DB 769 GVRGLKSGKSGEDGPGFKGDMGLKDRGEVQIGPRGXDPGPKGRAGTGDGPFS 828  
QY 61 GPPGEKGLGVPLGPGYGRQPKSGISGFPFGPANGKGGRTGPKGPRGGRGPTGPR 120  
DB 829 GQAGEKGLGVPLGPGYGRQPKSGISGFPFGPANGKGGRTGPKGPRGGRGPTGPR 888  
QY 121 GERPRGRTGPKGSGGDPAGPGRGPNQPGTFFPKGPPGPKGDKGLPGHP 180  
DB 889 GSRGARGPTGPKGSGGDPAGPGRGPNQPGTFFPKGPPGPKGDKGLPGHP 948  
QY 181 GORGET 186  
DB 949 GORGET 954

RESULT 5  
US-10-058-270A-122  
; Sequence 122, Application US/10058270A  
; Publication No. US20040029114A1  
; GENERAL INFORMATION:  
; APPLICANT: Mack, David H.  
; APPLICANT: Gish, Kurt C.  
; APPLICANT: Afar, Daniel  
; APPLICANT: Eos biotechnology, Inc.  
; TITLE OF INVENTION: Methods of Diagnosis of Breast Cancer, Compositions and  
; FILE REFERENCE: 018501-005210US  
; CURRENT APPLICATION NUMBER: US/10/058,270A  
; CURRENT FILING DATE: 2002-01-24  
; PRIOR APPLICATION NUMBER: US 60/263,965  
; PRIOR FILING DATE: 2001-01-24  
; PRIOR APPLICATION NUMBER: US 60/265,928  
; PRIOR FILING DATE: 2001-02-02  
; PRIOR APPLICATION NUMBER: US 09/829,472  
; PRIOR FILING DATE: 2001-04-09  
; PRIOR APPLICATION NUMBER: US 60/282,698  
; PRIOR FILING DATE: 2001-04-09  
; PRIOR APPLICATION NUMBER: US 60/288,590  
; PRIOR FILING DATE: 2001-05-04  
; PRIOR APPLICATION NUMBER: US 60/294,443  
; PRIOR FILING DATE: 2001-05-29  
; NUMBER OF SEQ ID NOS: 141  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 122  
; LENGTH: 1806  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (1)..(1806)  
; OTHER INFORMATION: Xaa = any amino acid  
US-10-058-270A-122

Query Match 86.2%; Score 930; DB 12; Length 1806;  
Best Local Similarity 86.0%; Pred. No. 1.8e-54;  
Matches 160; Conservative 5; Mismatches 21; Indels 0; Gaps 0;  
QY 1 GIRGLKGTGKSGDGGPGFKGDMGIKDRGEIGPRGPRGEDGEGPKGSGGNGDGPPL 60  
DB 769 GVRGLKSGKSGEDGPGFKGDMGLKDRGEVQIGPRGXDPGPKGRAGTGDGPFS 828  
QY 61 GPPGEKGLGVPLGPGYGRQPKSGISGFPFGPANGKGGRTGPKGPRGGRGPTGPR 120  
DB 829 GQAGEKGLGVPLGPGYGRQPKSGISGFPFGPANGKGGRTGPKGPRGGRGPTGPR 888  
QY 121 GERPRGRTGPKGSGGDPAGPGRGPNQPGTFFPKGPPGPKGDKGLPGHP 180



FILE REFERENCE: FP0402.3 CON  
CURRENT APPLICATION NUMBER: US/10/402,089  
CURRENT FILING DATE: 2003-03-26  
PRIOR APPLICATION NUMBER: US 09/709,700  
PRIOR FILING DATE: 2000-11-10  
NUMBER OF SEQ ID NOS: 72  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 6  
LENGTH: 1466  
TYPE: PRT  
ORGANISM: Sus scrofa  
US-10-402-089-6

Query Match 47.2%; Score 509.5; DB 15; Length 1466;  
Best Local Similarity 38.5%; Pred. No. 2.3e-26;  
Matches 109; Conservative 14; Mismatches 61; Indels 99; Gaps 6;  
1 GIRLKGKT-----KGEKEDGFPFGKDMGIKGRGEI---GPPGPRGED 42  
639 GLOGLPTGPPGNGKPGKGEAGAPGIPGKGDSDGAPGERGPPGAGGPPGPRGGA 698  
43 GPEGPKG-----RGGPN-----54  
699 GPPGEGKGAAGPPGPPGAGTGLQMPGERGPGPKDKGEPGSSGVDGAPGKD 758  
55 -----GDPGLGPPGKGLGVPLPGYP-----GRQPKGSIGPPGPPGAN 96  
759 GPRGPTGPIGPPGAGPGDKGSGAPGVGIAGPRGPGERBQGGPPGAGPPGAPGQN 818  
97 GEKGGRTGPKGPRGQ-----RGPTGPRGERGPRGITGKPGKGNSSGD 141  
819 GEPKAGGERGAPGKGEKGGPPGAAGGAGGPPGPPGVKGERSGPGGAGGPPGGR 878  
142 GPAGPPGERGPNQGTGTPGPKPGPPGPKDGLFGHPGQRG 184  
879 GPPGPPGSGNPGPPGSSGAPGKDGPPGPPGSGNAPGSPGISG 921  
US-10-402-072A-6

Query Match 47.2%; Score 509.5; DB 15; Length 1466;  
Best Local Similarity 38.5%; Pred. No. 2.3e-26;  
Matches 109; Conservative 14; Mismatches 61; Indels 99; Gaps 6;  
1 GIRLKGKT-----KGEKEDGFPFGKDMGIKGRGEI---GPPGPRGED 42  
639 GLOGLPTGPPGNGKPGKGEAGAPGIPGKGDSDGAPGERGPPGAGGPPGPRGGA 698  
43 GPEGPKG-----RGGPN-----54  
699 GPPGEGKGAAGPPGPPGAGTGLQMPGERGPGPKDKGEPGSSGVDGAPGKD 758  
55 -----GDPGLGPPGKGLGVPLPGYP-----GRQPKGSIGPPGPPGAN 96  
759 GPRGPTGPIGPPGAGPGDKGSGAPGVGIAGPRGPGERBQGGPPGAGPPGAPGQN 818  
97 GEKGGRTGPKGPRGQ-----RGPTGPRGERGPRGITGKPGKGNSSGD 141  
819 GEPKAGGERGAPGKGEKGGPPGAAGGAGGPPGPPGVKGERSGPGGAGGPPGGR 878  
142 GPAGPPGERGPNQGTGTPGPKPGPPGPKDGLFGHPGQRG 184  
879 GPPGPPGSGNPGPPGSSGAPGKDGPPGPPGSGNAPGSPGISG 921  
US-10-402-072A-4

Query Match 47.2%; Score 509.5; DB 15; Length 1466;  
Best Local Similarity 38.5%; Pred. No. 2.3e-26;  
Matches 109; Conservative 14; Mismatches 61; Indels 99; Gaps 6;  
1 GIRLKGKT-----KGEKEDGFPFGKDMGIKGRGEI---GPPGPRGED 42  
639 GLOGLPTGPPGNGKPGKGEAGAPGIPGKGDSDGAPGERGPPGAGGPPGPRGGA 698  
43 GPEGPKG-----RGGPN-----54  
699 GPPGEGKGAAGPPGPPGAGTGLQMPGERGPGPKDKGEPGSSGVDGAPGKD 758  
55 -----GDPGLGPPGKGLGVPLPGYP-----GRQPKGSIGPPGPPGAN 96  
759 GPRGPTGPIGPPGAGPGDKGSGAPGVGIAGPRGPGERBQGGPPGAGPPGAPGQN 818  
97 GEKGGRTGPKGPRGQ-----RGPTGPRGERGPRGITGKPGKGNSSGD 141  
819 GEPKAGGERGAPGKGEKGGPPGAAGGAGGPPGPPGVKGERSGPGGAGGPPGGR 878  
142 GPAGPPGERGPNQGTGTPGPKPGPPGPKDGLFGHPGQRG 184  
879 GPPGPPGSGNPGPPGSSGAPGKDGPPGPPGSGNAPGSPGISG 921  
US-10-402-072A-3

Query Match 47.2%; Score 509.5; DB 15; Length 1466;  
Best Local Similarity 38.5%; Pred. No. 2.3e-26;  
Matches 109; Conservative 14; Mismatches 61; Indels 99; Gaps 6;  
1 GIRLKGKT-----KGEKEDGFPFGKDMGIKGRGEI---GPPGPRGED 42  
639 GLOGLPTGPPGNGKPGKGEAGAPGIPGKGDSDGAPGERGPPGAGGPPGPRGGA 698  
43 GPEGPKG-----RGGPN-----54  
699 GPPGEGKGAAGPPGPPGAGTGLQMPGERGPGPKDKGEPGSSGVDGAPGKD 758  
55 -----GDPGLGPPGKGLGVPLPGYP-----GRQPKGSIGPPGPPGAN 96  
759 GPRGPTGPIGPPGAGPGDKGSGAPGVGIAGPRGPGERBQGGPPGAGPPGAPGQN 818  
97 GEKGGRTGPKGPRGQ-----RGPTGPRGERGPRGITGKPGKGNSSGD 141  
819 GEPKAGGERGAPGKGEKGGPPGAAGGAGGPPGPPGVKGERSGPGGAGGPPGGR 878  
142 GPAGPPGERGPNQGTGTPGPKPGPPGPKDGLFGHPGQRG 184  
879 GPPGPPGSGNPGPPGSSGAPGKDGPPGPPGSGNAPGSPGISG 921  
US-10-402-072A-2

; CURRENT FILING DATE: 2003-03-26  
 ; PRIOR APPLICATION NUMBER: US 09/709,700  
 ; PRIOR FILING DATE: 2000-11-10  
 ; NUMBER OF SEQ ID NOS: 72  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 12  
 ; LENGTH: 1466  
 ; TYPE: PRT  
 ; ORGANISM: Sus scrofa  
 US-10-402-089-12

Query Match 45.6%; Score 492.5; DB 15; Length 1466;  
 Best Local Similarity 41.9%; Pred. No. 3.2e-25;  
 Matches 104; Conservative 11; Mismatches 70; Indels 63; Gaps 5;

QY 1 GIRLKGTTGKGEDEGFPFGKDMGKIDRGEIGPPRGEDGPEGPKRGGPNGDGPGPL 60  
 DB 574 GQGVWGFPPGKNDGAPKNGRGGPGGFLPGPPKNGETGPGQPPGTPGGDKGDT 633  
 QY 61 GPPGEKGLGVPLP-----GYPGRQGPKGSGIGPPFGANGKXGGRGTPGKP----- 108  
 DB 634 GPPGQGLQGLPCTSPGKNGKPGEPGPKGAGAPGIPGKGDGSGAPGERGPPGAVGPS 693  
 QY 109 GPRGQRGTPGPRGERP-----RGITKPGPKGNS---GGDGPA 144  
 DB 694 GPRGGAGPPGEGKGPAGPPGPPGAAAGTPTGLQGMPPGERGSGGPGPKGDKGDPGSGAD 753  
 QY 145 GPPGERGPNQPGTGFPRGPKGPP-----GPPGKGLP 177  
 DB 754 GAFGKDGPRGTPGIPGPPGAPGQPDKGESGAPGLPGIAGPRGGGERGHEGPPGAGFP 813  
 QY 178 GHPGQGE 185  
 DB 814 GAPQNGE 821

RESULT 13  
 US-10-402-072A-12  
 ; Sequence 12, Application US/10402072A  
 ; Publication No. US20040018592A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bell, Marcum P.  
 ; APPLICANT: Neff, Thomas B.  
 ; APPLICANT: Polarek, James W.  
 ; APPLICANT: Seeley, Todd W.  
 ; TITLE OF INVENTION: BOVINE COLLAGENS AND GELATINS  
 ; FILE REFERENCE: EP0402.2 CON  
 ; CURRENT APPLICATION NUMBER: US/10/402, 072A  
 ; CURRENT FILING DATE: 2003-03-26  
 ; PRIOR APPLICATION NUMBER: US 09/709,700  
 ; PRIOR FILING DATE: 2000-11-10  
 ; NUMBER OF SEQ ID NOS: 72  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 12  
 ; LENGTH: 1466  
 ; TYPE: PRT  
 ; ORGANISM: Sus scrofa  
 US-10-402-072A-12

Query Match 45.6%; Score 492.5; DB 15; Length 1466;  
 Best Local Similarity 41.9%; Pred. No. 3.2e-25;  
 Matches 104; Conservative 11; Mismatches 70; Indels 63; Gaps 5;

QY 1 GIRLKGTTGKGEDEGFPFGKDMGKIDRGEIGPPRGEDGPEGPKRGGPNGDGPGPL 60  
 DB 574 GQGVWGFPPGKNDGAPKNGRGGPGGFLPGPPKNGETGPGQPPGTPGGDKGDT 633  
 QY 61 GPPGEKGLGVPLP-----GYPGRQGPKGSGIGFPFGANGKXGGRGTPGKP----- 108  
 DB 634 GPPGQGLQGLPCTSPGKNGKPGEPGPKGAGAPGIPGKGDGSGAPGERGPPGAVGPS 693  
 QY 109 GPRGQRGTPGPRGERP-----RGITKPGPKGNS---GGDGPA 144

Db 694 GPRGAGPPGEGGKGPAGPPPPGAAGTPTGLQMPGRCGSGGPGPKGDKGDPGSGAD 753  
 QY 145 GPPGERGPNQPGTGFPRGPKGPP-----GPPGKGLP 177  
 DB 754 GAPCKDGRGTPGIPGPPGAPGQPDKGESGAPGLPGIAGPRGGGERGHEGPPGAGFP 813  
 QY 178 GHPGQGE 185  
 DB 814 GAPQNGE 821

RESULT 14  
 US-10-177-293-70  
 ; Sequence 70, Application US/10177293  
 ; Publication No. US20030124128A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lillie, James  
 ; APPLICANT: Glatt, Karen  
 ; APPLICANT: Zhao, Xumei  
 ; APPLICANT: Gannavarpu, Manjula  
 ; APPLICANT: Kamatkar, Shubhangi  
 ; APPLICANT: Mertens, Maureen  
 ; APPLICANT: Myer, Vic  
 ; APPLICANT: Wang, Youzhen  
 ; APPLICANT: Xu, Yongyao  
 ; APPLICANT: Hoersch, Sebastian  
 ; APPLICANT: Monahan, John  
 ; APPLICANT: Meyers, Rachel E.  
 ; APPLICANT: Bast Jr., Robert C.  
 ; APPLICANT: Hortobagyi, Gabriel N.  
 ; APPLICANT: Pusztai, Lajos  
 ; APPLICANT: Meric, Funda  
 ; APPLICANT: Sahin, Aysegul  
 ; APPLICANT: Mills, Gordon B.  
 ; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,  
 ; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER  
 ; FILE REFERENCE: MRI-038  
 ; CURRENT APPLICATION NUMBER: US/10/177,293  
 ; CURRENT FILING DATE: 2002-06-21  
 ; PRIOR APPLICATION NUMBER: US 60/299,887  
 ; PRIOR FILING DATE: 2001-06-21  
 ; PRIOR APPLICATION NUMBER: US 60/301,572  
 ; PRIOR FILING DATE: 2001-06-27  
 ; PRIOR APPLICATION NUMBER: US 60/306,501  
 ; PRIOR FILING DATE: 2001-07-18  
 ; PRIOR APPLICATION NUMBER: US 60/325,002  
 ; PRIOR FILING DATE: 2001-09-25  
 ; PRIOR APPLICATION NUMBER: US 60/362,585  
 ; PRIOR FILING DATE: 2002-03-05  
 ; PRIOR APPLICATION NUMBER: US 60/xxx,xxx  
 ; PRIOR FILING DATE: 2002-05-14  
 ; NUMBER OF SEQ ID NOS: 506  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 70  
 ; LENGTH: 1496  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-177-293-70

Query Match 45.5%; Score 491; DB 14; Length 1496;  
 Best Local Similarity 51.4%; Pred. No. 4.1e-25;  
 Matches 95; Conservative 12; Mismatches 78; Indels 0; Gaps 0;  
 QY 1 GIRLKGTTGKGEDEGFPFGKDMGKIDRGEIGPPRGEDGPEGPKRGGPNGDGPGPL 60  
 DB 291 GAGLPLGLKHGRHKGLESGKGEVGA PGSGKEAGTGPWGAMGPIGPRGMFGERGRLGFPQ 350  
 QY 61 GPPGEKGLGVPLPGYPRGQPKGSGIGFPFGANGKXGGRGTPGKPGPRGQRTGPR 120  
 DB 351 GAPGQRGAHGMFKDGPMPGLGIPGSSGFPNGPMKGEAGTGA RGPGEQGRGETGPP 410  
 QY 121 GERGPRGITGKPGPKGNSGGDPAGPPGERGPNQGTPGFPKPGKPPGPKGDLPGHP 180



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 13:06:19 ; Search time 8.09953 Seconds  
(without alignments)

2208.970 Million cell updates/sec

Title: US-09-775-964-6

Perfect score: 1079

Sequence: 1 GIRGLKGTGKEKCEDGPPGF.....PPGPPGKDGLPQHPGQRGET 186

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1079	100.0	1838	1 CGHUIV	collagen alpha 1(V)
2	1071	99.3	1843	2 S18803	collagen alpha 1(V)
3	953	88.3	1024	2 S18251	collagen alpha 1(X)
4	930	86.2	1806	1 CGHUIE	collagen alpha 1(X)
5	873	80.9	1546	1 CGH02E	collagen alpha 2(X)
6	509.5	47.2	1049	1 CGB07S	collagen alpha 1(I)
7	501	46.4	1669	1 CGM34B	collagen alpha 1(I)
8	498.5	46.2	1691	1 S22917	collagen alpha 1(I)
9	497.5	46.1	888	2 S22971	collagen alpha 1(X)
10	494	45.8	674	2 S23297	collagen alpha 1(X)
11	491	45.5	1496	1 CGH02V	collagen alpha 2(V)
12	489.5	45.4	1464	2 S59856	collagen alpha 1(I)
13	488.5	45.3	1019	1 A32856	collagen alpha 1(V)
14	488.5	45.3	1492	2 A40333	collagen alpha 1(I)
15	486.5	45.1	1466	1 CGH07L	collagen alpha 1(I)
16	486	45.0	1669	1 CGH04B	collagen alpha 1(I)
17	485.5	45.0	674	2 S13301	collagen alpha 1(X)
18	484.5	44.9	1486	1 B40333	collagen alpha 1(I)
19	484	44.9	1532	2 A61262	collagen alpha 1(X)
20	483	44.8	964	1 CGCH2S	collagen alpha 2(I)
21	481.5	44.6	1146	2 A38587	collagen, cornea-s
22	480.5	44.5	920	2 A45748	collagen alpha 1(V)
23	479.5	44.4	471	2 A39024	collagen alpha 3(I)
24	479.5	44.4	671	1 CGRT1S	collagen alpha 1(I)
25	479.5	44.4	886	2 I50694	collagen alpha 1(I)
26	479.5	44.4	1418	2 T45467	collagen alpha 1(I)
27	479	44.4	402	1 CGB02S	collagen alpha 2(I)
28	479	44.4	488	2 A27353	collagen alpha 1(I)
29	478.5	44.3	1464	1 CGHUIS	collagen alpha 1(I)

#### RESULT 1

CGHUIV

collagen alpha 1(V) chain precursor - human

N;Alternate names: procollagen alpha 1(V) chain

C;Species: Homo sapiens (man)

C;Date: 22-Nov-1993 #sequence revision 03-Oct-1995 #text change 16-Jun-2000

C;Accession: S18802; S16024; A61142; S11303; S03978; S43642; S58665

R;GreenSpan, D.S.; Cheng, W.; Hoffman, G.G.

J. Biol. Chem. 266, 24727-24733, 1991

A;Title: The pro-alpha1(V) collagen chain. Complete primary structure, distribution of

A;Reference number: S18802; MUID:92105142; PMID:1722213

A;Accession: S18802

A;Molecule type: mRNA

A;Residues: 1-1838 <GRE>

A;Cross-references: GB:M76729; NID:gl89519; PIDN:AAAS9993.1; PID:gl89520

R;Takahara, K.; Sato, Y.; Okazawa, K.; Okamoto, N.; Noda, A.; Yaoi, Y.; Kato, I.

J. Biol. Chem. 266, 13124-13129, 1991

A;Title: Complete primary structure of human collagen alpha-1(V) chain.

A;Reference number: S16024; MUID:91302336; PMID:2071595

A;Accession: S16024

A;Molecule type: mRNA

A;Residues: 1-81, 'QU', '84-389, 'A', '391-676, 'K', '678-1294, 'PS', '1297, 'RS', '1300-1553, 'R', '1555

A;Cross-references: GB:D90279; NID:G219509; PIDN:BAAI4323.1; PID:G219510

A;Note: parts of this sequence were determined by protein sequencing

R;Yaoi, Y.; Hashimoto, K.; Takahara, K.; Kato, I.

Exp. Cell Res. 194, 180-185, 1991

A;Title: Insulin binds to type V collagen with retention of mitogenic activity.

A;Reference number: A61142; MUID:91224163; PMID:1709100

A;Accession: A61142

A;Molecule type: protein

A;Residues: 823-824, 'X', '826-842 <YAO>

A;Note: the residue designated 'X' is probably glycosylated hydroxylysine; this cyanogen

R;Yaoi, Y.; Hashimoto, K.; Koitabashi, H.; Takahara, K.; Ito, M.; Kato, I.

Biochim. Biophys. Acta 1035, 139-145, 1990

A;Title: Primary structure of the heparin-binding site of type V collagen.

A;Reference number: S11303; MUID:90366601; PMID:2203476

A;Accession: S11303

A;Molecule type: protein

A;Residues: 823-824, 'X', '826-848, 'I', '850-851, 'P', '853, 'PR', '856-893, 'D', '895-932, 'X', '934-951

A;Note: the residues designated 'X' are probably glycosylated hydroxylysine; this sequ

R;Seyer, J.M.; Kang, A.H.

Arch. Biochem. Biophys. 271, 120-129, 1989

A;Title: Covalent structure of collagen: amino acid sequence of three cyanogen bromide-

A;Reference number: S03978; MUID:89227189; PMID:2496661

A;Accession: S03978

A;Molecule type: protein

A;Residues: 621-640, 'G', '642-649, 'L', '651-662, 'E', '664-667, 'Q', '669-676, 'Q', '678-683, 'P', '685

<SE>

A;Note: there are a number of inconsistencies between the sequences in figures 6 and 7;

R;Moradi-Ameli, M.; Rousseau, J.C.; Kleman, J.P.; Champiaud, M.F.; Boutillon, M.M.; Be

Eur. J. Biochem. 221, 987-995, 1994

A;Title: Diversity in the processing events at the N-terminus of type-V collagen.





A;Title: Structural analysis of cross-linking domains in cartilage type XI collagen. Ins  
A;Reference number: A56978; MUID:95370194; PMID:7642541  
A;Accession: A56978  
A;Status: preliminary  
A;Molecule type: protein  
A;Residues: 206-229 <WUA>  
R;Niyibizi, C.; Eyre, D.R.  
Eur. J. Biochem. 224, 943-950, 1994  
A;Title: Structural characteristics of cross-linking sites in type V collagen of bone. C  
A;Reference number: S48210; MUID:95010086; PMID:7925418  
A;Accession: S65864  
A;Molecule type: protein  
A;Residues: 'X', 273-298 <NIY>  
C;Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology  
C;Keywords: hydroxyproline  
F;211,223/Modified site: 4-hydroxyproline (Pro) #status experimental

Query Match 88.3%; Score 953; DB 2; Length 1024;  
Best Local Similarity 88.2%; Pred. No. 3.2e-52;  
Matches 164; Conservative 4; Mismatches 18; Indels 0; Gaps 0;

Qy 1 GIRLKGTKGKEDGDFPGFKDMGIRKDRGEIGPPGPRGEDGPEGKGRGPNPNDPGL 60  
Db 536 GVRGLKSGKGEKEDGDFPGFKDMGLKDRGEVGVGPRGEDGPEGKGRAGPTGDPGP 595

Qy 61 GPPEKGLGVPLGPGVPGROGPKSGTGPGFPAGNKGKGRGTPPGKPRGQGRGTPGR 120  
Db 596 GQAGEKGLGVPLGPGVPGROGPKSGTGPGFPAGNKGKGRGTPPGKPRGQGRGTPGR 655

Qy 121 GERPRGITGKPGKNSGGDPAGPGERGPNPQGTGPPGKPGPPGKDGKGLPGHP 180  
Db 656 GSRGARGPTGKPGKNSGGDPAGPGERGPNPQGTGPPGKPGPPGKDGKGLPGHP 715

Qy 181 GQRGET 186  
Db 716 GQRGET 721

RESULT 4  
CGHUZE  
collagen alpha 1(XI) chain precursor - human  
N;Alternate names: procollagen alpha 1(XI) chain  
C;Species: Homo sapiens (man)  
C;Date: 31-Mar-1990 #sequence\_revision 03-Oct-1995 #text\_change 08-May-1998  
C;Accession: A35239; A31795  
R;Yoshioka, H.; Ramirez, F.  
J. Biol. Chem. 265, 6423-6426, 1990  
A;Title: Pro-alpha1(XI) collagen. Structure of the amino-terminal propeptide and expres  
A;Reference number: A35239; MUID:90202924; PMID:1690726  
A;Accession: A35239  
A;Molecule type: mRNA  
A;Residues: 1-558 <YOS>  
R;Bernard, M.; Yoshioka, H.; Rodriguez, E.; van der Rest, M.; Kimura, T.; Ninomiya, Y.;  
J. Biol. Chem. 263, 17159-17166, 1988  
A;Title: Cloning and sequencing of pro-alpha1(XI) collagen cDNA demonstrates that type X  
cartilaginous tissue.  
A;Reference number: A92689; MUID:89034222; PMID:3182841  
A;Accession: A31795  
A;Molecule type: DNA; mRNA  
A;Residues: 538-1806 <BER>  
A;Cross-references: GB:J04177  
A;Note: parts of this sequence were determined by protein sequencing  
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
ed and subsequently O-glycosylated.  
C;Genetics:  
A;Gene: GDB:COL11A1; COL16  
A;Cross-references: GDB:I20595; OMIM:120280  
A;Map position: 1p21-1p21  
A;Introns: 561/3; 579/3; 597/3; 615/3; 633/3; 648/3; 666/3; 681/3  
A;Note: the list of introns is incomplete  
C;Complex: type XI collagen may be a heterotrimer of two alpha 1(XI) chains and one alpha  
3(XI) chain (see PIR:CGHU6C), initially linked by disulfide bonds among their carboxyl

med with desmosome cross-links made from lysine and allysine residues  
C;Function:  
A;Description: structural component of extracellular fibrous polymer associated with ce  
A;Note: may play a role in controlling the lateral growth of collagen II fibrils  
C;Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology  
C;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyprol  
F;1-36/Domain: signal sequence #status predicted <SIG>  
F;35-260/Domain: PARP-like #status predicted <PARP>  
F;512-527/Region: amino-terminal propeptide #status predicted <PRO>  
F;512-1565/Product: collagen alpha 1(XI) chain #status predicted <MAT>  
F;528-1542/Region: helical  
F;1543-1565/Region: carboxyl-terminal nonhelical telopeptide  
F;1566-1806/Domain: carboxyl-terminal propeptide #status predicted <CTP>  
F;1583-1805/Domain: fibrillar collagen carboxyl-terminal homology <FCC>  
F;61-243,182-236/Disulfide bonds: #status predicted  
F;505/Modified site: allysine (Lys) #status predicted  
F;612,1452/Modified site: 5-hydroxylysine (Lys) #status predicted  
F;612,1452/Binding site: carbohydrate (Lys) (covalent) #status predicted

Query Match 86.2%; Score 930; DB 1; Length 1806;  
Best Local Similarity 86.0%; Pred. No. 1.4e-50;  
Matches 160; Conservative 5; Mismatches 21; Indels 0; Gaps 0;

Qy 1 GIRLKGTKGKEDGDFPGFKDMGIRKDRGEIGPPGPRGEDGPEGKGRGPNPNDPGL 60  
Db 769 GVRGLKSGKGEKEDGDFPGFKDMGLKDRGEVGVGPRGEDGPEGKGRAGPTGDPGP 828

Qy 61 GPPEKGLGVPLGPGVPGROGPKSGTGPPGFPAGNKGKGRGTPPGKPRGQGRGTPGR 120  
Db 829 GQAGEKGLGVPLGPGVPGROGPKSGTGPPGFPAGNKGKGRGTPPGKPRGQGRGTPGR 888

Qy 121 GERPRGITGKPGKNSGGDPAGPGERGPNPQGTGPPGKPGPPGKDGKGLPGHP 180  
Db 889 GSRGARGPTGKPGKNSGGDPAGPGERGPNPQGTGPPGKPGPPGKDGKGLPGHP 948

Qy 181 GQRGET 186  
Db 949 GQRGET 954

RESULT 5  
CGHUZE  
collagen alpha 2(XI) chain precursor - human (fragment)  
N;Alternate names: procollagen alpha 2(XI) chain  
N;Contents: proline/arginine-rich protein (PARP)  
C;Species: Homo sapiens (man)  
C;Date: 07-Jun-1990 #sequence\_revision 03-Oct-1995 #text\_change 22-Jun-1999  
C;Accession: S34790; A32645  
R;Zhidkova, N.I.; Brewton, R.G.; Mayne, R.  
FEBS Lett. 326, 25-28, 1993  
A;Title: Molecular cloning of PARP (proline/arginine-rich protein) from human cartilage  
chain.  
A;Reference number: S34790; MUID:93314796; PMID:8325374  
A;Accession: S34790  
A;Molecule type: mRNA  
A;Residues: 1-663 <ZHI>  
A;Cross-references: EMBL:L18987; NID:G306439; PIDN:AAA35498.1; PID:G306440  
R;Kimura, T.; Cheah, K.S.E.; Chan, S.D.H.; Lui, V.C.H.; Mattei, M.G.; van der Rest, M.;  
J. Biol. Chem. 264, 13910-13916, 1989  
A;Title: The human alpha2(XI) collagen (COL11A2) chain. Molecular cloning of cDNA and g  
A;Reference number: A32645; MUID:89340485; PMID:2760050  
A;Accession: A32645  
A;Molecule type: DNA; mRNA  
A;Residues: 586-1546 <KIM>  
A;Cross-references: GB:J04974; NID:G180714; PIDN:AAA52034.1; PID:G180715  
A;Note: parts of this sequence were determined by protein sequencing  
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
ed and subsequently O-glycosylated.  
C;Genetics:  
A;Gene: GDB:COL11A2  
A;Cross-references: GDB:I19788; OMIM:120290  
A;Map position: 6p21.3-6p21.3





QY 1 GIRGLKGTKEGEDGDPGFKGDMGI---KGRGETGPPRGEDG-----PEGPKG 49  
A;Reference number: A34850; MUID:90160375; PMID:1689491  
A;Accession: A34850  
A;Molecule type: mRNA  
Db 1177 GPGSGKDGSGKEGVFPGLAGSPGIPGVKGQGMGPPGQPGQLPCTPGHPVEGPKG 1236  
A;Cross-references: EMBL:M31115; NID:g180824; PIDN:AAA52045.1; PID:g180825  
R;Zhou, J.; Hostikka, S.L.; Chow, L.T.; Tryggvason, K.  
QY 50 RGGPNGD-----PGPLGPP-----GEKKGVLVPLGVPGRQPGKSGIGPPGP-- 92  
A;Title: Characterization of the 3' half of the human type IV collagen alpha-5 gene that  
Db 1237 DRGPGQPGPLGHPGPMGPPGPGINGPKDGKGNQWPGAPGVPGPKGDPGFGMGPGIGG 1296  
A;Reference number: A37969; MUID:91169491; PMID:2004755  
A;Accession: S18850  
QY 93 -PCANGEKGRGTGPGKPGRQGRGPTGPRGSGRCITGKPGKGNSGDGPAGP----- 146  
A;Molecule type: DNA  
Db 1297 SPGITSGKGDMLPGVPFGQKGLPGLQGVKGDQGVPGPKLQGLGPPGPPGYDVIK 1356  
A;Cross-references: EMBL:M63455; EMBL:M63456; EMBL:M63457; EMBL:M63458; EMBL:M63459; EMBL:M63470; EMBL:M63471; EMBL:M63472; EMBL:M63473; NID:g177922; PIDN:AAA51558.1; PID:8  
R;Guo, C.; Van Damme, B.; Van Damme-Lombaerts, R.; Van den Berghe, H.; Cassiman, J.J.; M  
QY 147 --PGERGPNPGQPTGPPGPKGP-----PGPPKGDLPGHPQORGET 186  
A;Title: Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex  
Db 1357 GEPGLGPPGPKLGLQFPGPKGOOGVTSVGLPGPVGFDGAPQKQGET 1410  
A;Reference number: I56971; MUID:94133540; PMID:8301933  
A;Accession: I56971  
A;Status: translated from GB/EMBL/DBDJ  
RESULT 8  
S22917  
collagen alpha 5(IV) chain precursor, renal splice form - human  
N;Alternate names: procollagen alpha 5(IV) chain  
N;Contains: collagen alpha 5(IV) chain precursor, leukocyte splice form  
C;Species: Homo sapiens (man)  
C;Date: 30-Sep-1993 #sequence revision 27-Feb-1997 #text change 21-Jul-2000  
C;Accession: S22917; A54365; A57079; A37122; I54317; A34850; S18850; I56971; I76598; A35  
R;Zhou, J.; Hertz, J.M.; Leinonen, A.; Tryggvason, K.  
J. Biol. Chem. 267, 12475-12481, 1992  
A;Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and identi  
n Alport syndrome patient.  
A;Reference number: S22917; MUID:92316923; PMID:1352287  
A;Accession: S22917  
A;Molecule type: mRNA  
A;Residues: 1-967 <ZH>  
A;Cross-references: GB:M90464; NID:g180826; PIDN:AAA52046.1; PID:9553234  
R;Zhou, J.; Leinonen, A.; Tryggvason, K.  
J. Biol. Chem. 269, 6608-6614, 1994  
A;Title: Structure of the human type IV collagen COL4A5 gene.  
A;Reference number: A54365; MUID:94165049; PMID:8120014  
A;Accession: A54365  
A;Molecule type: DNA  
A;Residues: 1-922 <ZH2>  
A;Cross-references: GB:U04470; NID:g463378; GB:U04520; NID:g463428; PIDN:AAAC27816.1; PID  
R;Zhou, J.; Mochizuki, T.; Smeets, H.; Antignac, C.; Laurila, P.; de Paepe, A.; Tryggva  
Science 261, 1167-1169, 1993  
A;Title: Deletion of the paired alphas(IV) and alpha6(IV) collagen genes in inherited sm  
A;Reference number: A57079; MUID:93361972; PMID:8356449  
A;Accession: A57079  
A;Molecule type: DNA  
A;Residues: 1-27 <ZH4>  
A;Cross-references: GB:Z37153; NID:g587203; PIDN:CAA85512.1; PID:g587204  
R;Pihlajaniemi, T.; Pohjola, E.R.; Myers, J.C.  
J. Biol. Chem. 265, 13758-13766, 1990  
A;Title: Complete primary structure of the triple-helical region and the carboxyl-termin  
A;Reference number: A37122; MUID:90337990; PMID:2380186  
A;Accession: A37122  
A;Molecule type: mRNA  
A;Residues: 84-439, 'GS', 442-624, 'LALO', 629-666, 'ER', 669-887, 'R', 889-1264, 1271-1691 <PIH>  
A;Cross-references: GB:J05558; EMBL:M58526; NID:g1314209  
A;Note: submitted to the EMBL Data Library, February 1991  
A;Note: the authors translated the codon GCC for residue 115 as Val  
R;Renieri, A.; Seri, M.; Myers, J.C.; Pihlajaniemi, T.; Massella, L.; Rizzoni, G.; De Ma  
Hum. Mol. Genet. 1, 127-129, 1992  
A;Title: De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in  
A;Reference number: I54317; MUID:93244772; PMID:1363780  
A;Accession: I54317  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: mRNA  
A;Residues: 313-324, 'E', 326-330 <REN>  
A;Cross-references: GB:S59334; NID:g299946; PIDN:AAD13909.1; PID:g4261609  
R;Hostikka, S.L.; Eddy, R.L.; Byers, M.G.; Hoeyhtvae, M.; Shows, T.B.; Tryggvason, K.  
Proc. Natl. Acad. Sci. U.S.A. 87, 1606-1610, 1990  
A;Title: Identification of a distinct type IV collagen alpha chain with restricted kidney

A;Reference number: A34850; MUID:90160375; PMID:1689491  
A;Accession: A34850  
A;Molecule type: mRNA  
A;Residues: 914-1264, 1271-1691 <HOS>  
A;Cross-references: EMBL:M31115; NID:g180824; PIDN:AAA52045.1; PID:g180825  
R;Zhou, J.; Hostikka, S.L.; Chow, L.T.; Tryggvason, K.  
A;Title: Characterization of the 3' half of the human type IV collagen alpha-5 gene that  
A;Reference number: A37969; MUID:91169491; PMID:2004755  
A;Accession: S18850  
A;Molecule type: DNA  
A;Residues: 924-1264, 1271-1691 <ZH3>  
A;Cross-references: EMBL:M63455; EMBL:M63456; EMBL:M63457; EMBL:M63458; EMBL:M63459; EMBL:M63470; EMBL:M63471; EMBL:M63472; EMBL:M63473; NID:g177922; PIDN:AAA51558.1; PID:8  
R;Guo, C.; Van Damme, B.; Van Damme-Lombaerts, R.; Van den Berghe, H.; Cassiman, J.J.; M  
Kidney Int. 44, 1316-1321, 1993  
A;Title: Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex  
A;Reference number: I56971; MUID:94133540; PMID:8301933  
A;Accession: I56971  
A;Status: translated from GB/EMBL/DBDJ  
A;Molecule type: mRNA  
A;Residues: 1258-1276 <GUO1>  
A;Cross-references: GB:S69168; NID:g545095; PIDN:AAAC60612.1; PID:g545096  
A;Note: kidney splice form  
A;Accession: I76598  
A;Status: translated from GB/EMBL/DBDJ  
A;Molecule type: mRNA  
A;Residues: 1284-1291, 'TFLGVLACLV' <GUO2>  
A;Cross-references: GB:S69169; NID:g545097; PIDN:AAAC60613.1; PID:g545098  
A;Note: frameshift mutation in patient with Alport syndrome  
R;Myers, J.C.; Jones, T.A.; Pohjola, E.R.; Kadri, A.S.; Goddard, A.D.; Sheer, D.; Sc  
Am. J. Hum. Genet. 46, 1024-1033, 1990  
A;Title: Molecular cloning of alphas(IV) collagen and assignment of the gene to the regi  
A;Reference number: A35335; MUID:90252791; PMID:2339699  
A;Accession: A35335  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 1448-1477 <MYE>  
R;Nakazato, H.; Hattori, S.; Ushijima, T.; Matsumura, T.; Koitabashi, Y.; Takada, T.; Yos  
Kidney Int. 46, 1307-1314, 1994  
A;Title: Mutations in the COL4A5 gene in Alport syndrome: a possible mutation in primord  
A;Reference number: I56975; MUID:95156893; PMID:7853788  
A;Accession: I56975  
A;Status: translated from GB/EMBL/DBDJ  
A;Molecule type: DNA  
A;Residues: 1595-1602 <NAK>  
A;Cross-references: GB:S75903; NID:g913882; PIDN:AAB33374.1; PID:g913883  
A;Note: premature termination mutation from a patient with Alport syndrome; one other mu  
R;Lemmink, H.H.; Schroeder, C.H.; Brummer, H.G.; Nelen, M.R.; Zhou, J.; Tryggvason, K.;  
Genomics 17, 485-489, 1993  
A;Title: Identification of four novel mutations in the COL4A5 gene of patients with Alpo  
A;Reference number: I54188; MUID:94010948; PMID:8406498  
A;Accession: I54188  
A;Status: translated from GB/EMBL/DBDJ  
A;Molecule type: DNA  
A;Residues: 1604-1607, 'VHDAYKC' <LEM>  
A;Cross-references: GB:S65767; NID:g425563; PIDN:AAD13967.1; PID:g4261667  
A;Note: frameshift mutation from a patient with Alport syndrome; five other mutations a  
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
ed and subsequently O-glycosylated.  
C;Genetics:  
A;Gene: GDB:COL4A5; ATS  
A;Cross-references: GDB:120596; OMIM:303630  
A;Map position: Xq22-Xq22  
A;Intons: 97/3; 47/3; 77/3; 92/3; 107/3; 128/3; 146/3; 155/3; 182/3; 203/3; 215/3; 229/  
/3; 799/1; 837/1; 893/1; 923/1; 973/1; 1006/1; 1036/1; 1082/3; 1125/1; 1152/1; 1185/1; 1  
A;Note: the alpha 5(IV) and alpha 6(IV) chain genes are encoded on opposite strands with  
C;Complex: this minor type IV collagen is thought to form a heterotrimer of two alpha 5(  
mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric  
er associations in the interrupted helical domain (with disulfide and desmosine cross-li  
C;Function:  
A;Description: minor structural component of extracellular basement membrane  
C;Superfamily: collagen alpha 1(IV) chain

C;Keywords: Alport syndrome; basement membrane; coiled coil; extracellular matrix; glyco  
F;1-26/Domain: signal sequence #status predicted <SIG>  
F;27-1691/Product: collagen alpha 5(IV) chain, renal splice form #status predicted <MATR  
F;27-1264,1271-1691/Product: collagen alpha 5(IV) chain, leukocyte splice form #status p  
F;27-41/Domain: amino-terminal nonhelical, NC2 #status predicted <NC2>  
F;42-1462/Region: interrupted helical  
F;1463-1691/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>  
F;1473-1573/Domain: collagen IV carboxyl-terminal repeat <CTR>  
F;1583-1687/Domain: collagen IV carboxyl-terminal repeat <CTR>  
F;29,32,38,40,124,451,481,484/Disulfide bonds: interchain #status predicted  
F;15/Binding site: carboxylate (Asn) (covalent) #status predicted  
F;1482-1570,1515-1573/Disulfide bonds: (or 1482-1573, 1515-1570) #status predicted  
F;1527-1533,1638-1644/Disulfide bonds: #status predicted  
F;1592-1684,1626-1687/Disulfide bonds: (or 1592-1687, 1626-1684) #status predicted

Query Match 46.2%; Score 498.5; DB 1; Length 1691;  
Best Local Similarity 46.0%; Pred. No. 6.6e-24;  
Matches 108; Conservative 13; Mismatches 63; Indels 51; Gaps 7;

QY 1 GIRGLKTKGKEDGPPGKDMGKIGDRGE-----IGPPGRGEDGP--GGPKG 49  
Db 1193 GPPLGLSLGQKGDGGLPGIPGNPLGPKGEPFGHGFPGVQGGPPGPGSPGPALEGPKG 1252

QY 50 RCGPNGB-----PGLQPP-----GEKGLGVPLPGVP--GRGPKG 85  
Db 1253 NPFGQPPGPRPTGFGGLFQPEPPGLPGNGIKGKGNPQPGPLGLKGDQGGPPG 1312

QY 86 SIGFPPGPGANGKGRGTGPKPRGQRGTGPRGRGPR---GITGKPGP----- 134  
Db 1313 LQGNPGRPLGNKMGDPLGPGVGFPGCMKGPSGVPSAGPEGEPGLIGPPGLPGCPSG 1372

QY 135 -----KNSGGDGPAGPPGRGRNGPPGPGTGFPGPKGPPGPKOGLPHGPGQG 184  
Db 1373 QSIIRKGDAGPPGTPGQGLKGLPGPGQGLGPGTGPDPGRNGLPGPDGAGG 1427

RESULT 9  
S28791  
collagen alpha 1(XI) chain - chicken (fragment)  
C;Species: Gallus gallus (chicken)  
C;Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 13-Aug-1999  
C;Accession: S28791  
R;Nah, H.D.; Barenbaum, M.; Upholt, W.B.  
J. Biol. Chem. 267, 22581-22586, 1992  
A;Title: The chicken alpha 1(XI) collagen gene is widely expressed in embryonic tissues.  
A;Reference number: S28791; MUID:93054557; PMID:1429607  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-888 <NAH>  
A;Cross-references: EMBL:M88593; NID:g211619; PIDN:AAA48707.1; PID:g211620  
C;Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology  
F;665-887/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 46.1%; Score 497.5; DB 2; Length 888;  
Best Local Similarity 48.8%; Pred. No. 4.5e-24;  
Matches 103; Conservative 12; Mismatches 69; Indels 27; Gaps 6;

QY 1 GIRGLKTKGKEDGPPGKDMGKIGDRGEITGPPGRGEDGPEGKGRGGP--NGDP 57  
Db 10 GFGPGKPPGPPGKGLGPHGQRGTGFGKGTGPPGPGVGVGPGTGTGIGRGRHP 69

QY 58 GPLGPFGEKGLGV-----PGLPGYGRGPKXSGISGFP---GPPGANGS---KGRG 102  
Db 70 GPPGPPGEQQLPGAAGKEGAKGDPGQGIPIKGDPAGLRGFPGERGLPGAQGPAGLKGGE 129

QY 103 GTPGKPGPRQRGTGPRGRGRGTGPKPKNGSGGDPGAPGPRGPNGB-----Q 156  
Db 130 GPQGPFGVPSGERGAAGTAGTIGLPGRPGQPPGPGPAGEKAGPEKGPQGPAGRQGVQ 189

QY 157 GPTGFGPKGPPGPKGDKLP---GHPGQG 184  
Db 190 GPVGLPGPAGPSGPDGDKGIBGEPQKG 220

## RESULT 10

S23297

collagen alpha 1(X) chain precursor - chicken

N;Alternate names: type X collagen

C;Species: Gallus gallus (chicken)

C;Date: 07-Oct-1994 #sequence\_revision 10-Nov-1995 #text\_change 13-Aug-1999

C;Accession: S23297; A31896; S65594; S77711; I50218

R;Ninomiya, Y.; Castagnola, P.; Gerecke, D.; Gordon, M.K.; Jacenko, O.; LuValle, P.; M

maguchi, N.; Olsen, B.R.

In Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 79-114. Academic P

A;Title: The molecular biology of collagens with short triple-helical domains.

A;Reference number: S22243

A;Accession: S23297

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-674 &lt;NIN&gt;

R;LuValle, P.; Ninomiya, Y.; Rosenblum, N.D.; Olsen, B.R.

J. Biol. Chem. 263, 18378-18385, 1988

A;Title: The type X collagen gene. Introns sequences split the 5'-untranslated region an

A;Reference number: A31896; MUID:89054019; PMID:2461368

A;Accession: A31896

A;Molecule type: mRNA

A;Residues: 1-75 &lt;LUV&gt;

R;Ninomiya, Y.; Gordon, M.; van der Rest, M.; Schmid, T.; Linsemayer, T.; Olsen, B.R.

J. Biol. Chem. 261, 5041-5050, 1986

A;Title: The developmentally regulated type X collagen gene contains a long open readin

A;Reference number: I50218; MUID:86168227; PMID:3082876

A;Accession: S65594

A;Molecule type: DNA

A;Residues: 'T', '9', 'D', '11-12, 'EDMKLYLTFTM', '30-31, 'TCKSGRAFTTMYLONVMADLVSSHT', '48-89, 'L'

629, 'PQAVLSLSWRTIKSGCOIQNMVSIPLNMIFLLSQVSVLLKNNIPLTMS', &lt;NIN1&gt;

A;Cross-references: EMBL:M13496; NID:g211699; PIDN:AAA48736.1; PID:g211700

A;Accession: S77711

A;Molecule type: protein

A;Residues: 104-112, 'X', '114-117,453-466 &lt;NIN2&gt;

C;Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology

C;Keywords: coiled coil; extracellular matrix; glycoprotein; homotrimer; hydroxyproline

F;1-18/Domain: signal sequence #status predicted &lt;SIG&gt;

F;147-673/Domain: complement C1q carboxyl-terminal homology &lt;CIQ&gt;

F;453,456/Modified site: hydroxyproline (Pro) #status experimental

F;611/Binding site: carboxylate (Asn) (covalent) #status predicted





A;Reference number: S59856; MUID:95011609; PMID:7926795  
A;Accession: S59856  
A;Molecule type: DNA  
A;Residues: 1-1464 <TOM>  
A;Cross-references: EMBL:X52046  
R;Toman, D.  
submitted to the EMBL Data Library, November 1994  
A;Reference number: S62120  
A;Accession: S62120  
A;Molecule type: DNA  
A;Residues: 1-866, 'G', 868-1464 <TOM>  
A;Cross-references: EMBL:X52046; NID:9575321; PIDN:CAA36279.1; PID:9575322  
R;Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.  
Biochim. Biophys. Acta 1089, 241-243, 1991  
A;Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.  
A;Reference number: S16176; MUID:91274355; PMID:2054384  
A;Accession: S16173  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1442-1464 <MET>  
A;Cross-references: EMBL:X57983; NID:950476; PIDN:CAA1048.1; PID:950477  
C;Genetics:  
A;Introns: 29/1; 95/3; 112/3; 150/3; 175/3; 193/3; 211/3; 229/3; 247/3; 265/3; 283/3; 295/3; 313/3; 331/3; 349/3; 367/3; 385/3; 403/3; 421/3; 439/3; 457/3; 475/3; 493/3; 511/3; 529/3; 547/3; 565/3; 583/3; 601/3; 619/3; 637/3; 655/3; 673/3; 691/3; 709/3; 727/3; 745/3; 763/3; 781/3; 799/3; 817/3; 835/3; 853/3; 871/3; 889/3; 907/3; 925/3; 943/3; 961/3; 979/3; 997/3; 1015/3; 1033/3; 1051/3; 1069/3; 1087/3; 1105/3; 1123/3; 1141/3; 1159/3; 1177/3; 1195/3; 1213/3; 1231/3; 1249/3; 1267/3; 1285/3; 1303/3; 1321/3; 1339/3; 1357/3; 1375/3; 1393/3; 1411/3; 1429/3; 1447/3; 1465/3; 1483/3; 1501/3; 1519/3; 1537/3; 1555/3; 1573/3; 1591/3; 1609/3; 1627/3; 1645/3; 1663/3; 1681/3; 1699/3; 1717/3; 1735/3; 1753/3; 1771/3; 1789/3; 1807/3; 1825/3; 1843/3; 1861/3; 1879/3; 1897/3; 1915/3; 1933/3; 1951/3; 1969/3; 1987/3; 2005/3; 2023/3; 2041/3; 2059/3; 2077/3; 2095/3; 2113/3; 2131/3; 2149/3; 2167/3; 2185/3; 2203/3; 2221/3; 2239/3; 2257/3; 2275/3; 2293/3; 2311/3; 2329/3; 2347/3; 2365/3; 2383/3; 2401/3; 2419/3; 2437/3; 2455/3; 2473/3; 2491/3; 2509/3; 2527/3; 2545/3; 2563/3; 2581/3; 2599/3; 2617/3; 2635/3; 2653/3; 2671/3; 2689/3; 2707/3; 2725/3; 2743/3; 2761/3; 2779/3; 2797/3; 2815/3; 2833/3; 2851/3; 2869/3; 2887/3; 2905/3; 2923/3; 2941/3; 2959/3; 2977/3; 2995/3; 3013/3; 3031/3; 3049/3; 3067/3; 3085/3; 3103/3; 3121/3; 3139/3; 3157/3; 3175/3; 3193/3; 3211/3; 3229/3; 3247/3; 3265/3; 3283/3; 3301/3; 3319/3; 3337/3; 3355/3; 3373/3; 3391/3; 3409/3; 3427/3; 3445/3; 3463/3; 3481/3; 3499/3; 3517/3; 3535/3; 3553/3; 3571/3; 3589/3; 3607/3; 3625/3; 3643/3; 3661/3; 3679/3; 3697/3; 3715/3; 3733/3; 3751/3; 3769/3; 3787/3; 3805/3; 3823/3; 3841/3; 3859/3; 3877/3; 3895/3; 3913/3; 3931/3; 3949/3; 3967/3; 3985/3; 4003/3; 4021/3; 4039/3; 4057/3; 4075/3; 4093/3; 4111/3; 4129/3; 4147/3; 4165/3; 4183/3; 4201/3; 4219/3; 4237/3; 4255/3; 4273/3; 4291/3; 4309/3; 4327/3; 4345/3; 4363/3; 4381/3; 4399/3; 4417/3; 4435/3; 4453/3; 4471/3; 4489/3; 4507/3; 4525/3; 4543/3; 4561/3; 4579/3; 4597/3; 4615/3; 4633/3; 4651/3; 4669/3; 4687/3; 4705/3; 4723/3; 4741/3; 4759/3; 4777/3; 4795/3; 4813/3; 4831/3; 4849/3; 4867/3; 4885/3; 4903/3; 4921/3; 4939/3; 4957/3; 4975/3; 4993/3; 5011/3; 5029/3; 5047/3; 5065/3; 5083/3; 5101/3; 5119/3; 5137/3; 5155/3; 5173/3; 5191/3; 5209/3; 5227/3; 5245/3; 5263/3; 5281/3; 5299/3; 5317/3; 5335/3; 5353/3; 5371/3; 5389/3; 5407/3; 5425/3; 5443/3; 5461/3; 5479/3; 5497/3; 5515/3; 5533/3; 5551/3; 5569/3; 5587/3; 5605/3; 5623/3; 5641/3; 5659/3; 5677/3; 5695/3; 5713/3; 5731/3; 5749/3; 5767/3; 5785/3; 5803/3; 5821/3; 5839/3; 5857/3; 5875/3; 5893/3; 5911/3; 5929/3; 5947/3; 5965/3; 5983/3; 6001/3; 6019/3; 6037/3; 6055/3; 6073/3; 6091/3; 6109/3; 6127/3; 6145/3; 6163/3; 6181/3; 6199/3; 6217/3; 6235/3; 6253/3; 6271/3; 6289/3; 6307/3; 6325/3; 6343/3; 6361/3; 6379/3; 6397/3; 6415/3; 6433/3; 6451/3; 6469/3; 6487/3; 6505/3; 6523/3; 6541/3; 6559/3; 6577/3; 6595/3; 6613/3; 6631/3; 6649/3; 6667/3; 6685/3; 6703/3; 6721/3; 6739/3; 6757/3; 6775/3; 6793/3; 6811/3; 6829/3; 6847/3; 6865/3; 6883/3; 6901/3; 6919/3; 6937/3; 6955/3; 6973/3; 6991/3; 7009/3; 7027/3; 7045/3; 7063/3; 7081/3; 7099/3; 7117/3; 7135/3; 7153/3; 7171/3; 7189/3; 7207/3; 7225/3; 7243/3; 7261/3; 7279/3; 7297/3; 7315/3; 7333/3; 7351/3; 7369/3; 7387/3; 7405/3; 7423/3; 7441/3; 7459/3; 7477/3; 7495/3; 7513/3; 7531/3; 7549/3; 7567/3; 7585/3; 7603/3; 7621/3; 7639/3; 7657/3; 7675/3; 7693/3; 7711/3; 7729/3; 7747/3; 7765/3; 7783/3; 7801/3; 7819/3; 7837/3; 7855/3; 7873/3; 7891/3; 7909/3; 7927/3; 7945/3; 7963/3; 7981/3; 7999/3; 8017/3; 8035/3; 8053/3; 8071/3; 8089/3; 8107/3; 8125/3; 8143/3; 8161/3; 8179/3; 8197/3; 8215/3; 8233/3; 8251/3; 8269/3; 8287/3; 8305/3; 8323/3; 8341/3; 8359/3; 8377/3; 8395/3; 8413/3; 8431/3; 8449/3; 8467/3; 8485/3; 8503/3; 8521/3; 8539/3; 8557/3; 8575/3; 8593/3; 8611/3; 8629/3; 8647/3; 8665/3; 8683/3; 8701/3; 8719/3; 8737/3; 8755/3; 8773/3; 8791/3; 8809/3; 8827/3; 8845/3; 8863/3; 8881/3; 8899/3; 8917/3; 8935/3; 8953/3; 8971/3; 8989/3; 9007/3; 9025/3; 9043/3; 9061/3; 9079/3; 9097/3; 9115/3; 9133/3; 9151/3; 9169/3; 9187/3; 9205/3; 9223/3; 9241/3; 9259/3; 9277/3; 9295/3; 9313/3; 9331/3; 9349/3; 9367/3; 9385/3; 9403/3; 9421/3; 9439/3; 9457/3; 9475/3; 9493/3; 9511/3; 9529/3; 9547/3; 9565/3; 9583/3; 9601/3; 9619/3; 9637/3; 9655/3; 9673/3; 9691/3; 9709/3; 9727/3; 9745/3; 9763/3; 9781/3; 9799/3; 9817/3; 9835/3; 9853/3; 9871/3; 9889/3; 9907/3; 9925/3; 9943/3; 9961/3; 9979/3; 9997/3; 10015/3; 10033/3; 10051/3; 10069/3; 10087/3; 10105/3; 10123/3; 10141/3; 10159/3; 10177/3; 10195/3; 10213/3; 10231/3; 10249/3; 10267/3; 10285/3; 10303/3; 10321/3; 10339/3; 10357/3; 10375/3; 10393/3; 10411/3; 10429/3; 10447/3; 10465/3; 10483/3; 10501/3; 10519/3; 10537/3; 10555/3; 10573/3; 10591/3; 10609/3; 10627/3; 10645/3; 10663/3; 10681/3; 10699/3; 10717/3; 10735/3; 10753/3; 10771/3; 10789/3; 10807/3; 10825/3; 10843/3; 10861/3; 10879/3; 10897/3; 10915/3; 10933/3; 10951/3; 10969/3; 10987/3; 11005/3; 11023/3; 11041/3; 11059/3; 11077/3; 11095/3; 11113/3; 11131/3; 11149/3; 11167/3; 11185/3; 11203/3; 11221/3; 11239/3; 11257/3; 11275/3; 11293/3; 11311/3; 11329/3; 11347/3; 11365/3; 11383/3; 11401/3; 11419/3; 11437/3; 11455/3; 11473/3; 11491/3; 11509/3; 11527/3; 11545/3; 11563/3; 11581/3; 11599/3; 11617/3; 11635/3; 11653/3; 11671/3; 11689/3; 11707/3; 11725/3; 11743/3; 11761/3; 11779/3; 11797/3; 11815/3; 11833/3; 11851/3; 11869/3; 11887/3; 11905/3; 11923/3; 11941/3; 11959/3; 11977/3; 11995/3; 12013/3; 12031/3; 12049/3; 12067/3; 12085/3; 12103/3; 12121/3; 12139/3; 12157/3; 12175/3; 12193/3; 12211/3; 12229/3; 12247/3; 12265/3; 12283/3; 12301/3; 12319/3; 12337/3; 12355/3; 12373/3; 12391/3; 12409/3; 12427/3; 12445/3; 12463/3; 12481/3; 12499/3; 12517/3; 12535/3; 12553/3; 12571/3; 12589/3; 12607/3; 12625/3; 12643/3; 12661/3; 12679/3; 12697/3; 12715/3; 12733/3; 12751/3; 12769/3; 12787/3; 12805/3; 12823/3; 12841/3; 12859/3; 12877/3; 12895/3; 12913/3; 12931/3; 12949/3; 12967/3; 12985/3; 13003/3; 13021/3; 13039/3; 13057/3; 13075/3; 13093/3; 13111/3; 13129/3; 13147/3; 13165/3; 13183/3; 13201/3; 13219/3; 13237/3; 13255/3; 13273/3; 13291/3; 13309/3; 13327/3; 13345/3; 13363/3; 13381/3; 13399/3; 13417/3; 13435/3; 13453/3; 13471/3; 13489/3; 13507/3; 13525/3; 13543/3; 13561/3; 13579/3; 13597/3; 13615/3; 13633/3; 13651/3; 13669/3; 13687/3; 13705/3; 13723/3; 13741/3; 13759/3; 13777/3; 13795/3; 13813/3; 13831/3; 13849/3; 13867/3; 13885/3; 13903/3; 13921/3; 13939/3; 13957/3; 13975/3; 13993/3; 14011/3; 14029/3; 14047/3; 14065/3; 14083/3; 14101/3; 14119/3; 14137/3; 14155/3; 14173/3; 14191/3; 14209/3; 14227/3; 14245/3; 14263/3; 14281/3; 14299/3; 14317/3; 14335/3; 14353/3; 14371/3; 14389/3; 14407/3; 14425/3; 14443/3; 14461/3; 14479/3; 14497/3; 14515/3; 14533/3; 14551/3; 14569/3; 14587/3; 14605/3; 14623/3; 14641/3; 14659/3; 14677/3; 14695/3; 14713/3; 14731/3; 14749/3; 14767/3; 14785/3; 14803/3; 14821/3; 14839/3; 14857/3; 14875/3; 14893/3; 14911/3; 14929/3; 14947/3; 14965/3; 14983/3; 15001/3; 15019/3; 15037/3; 15055/3; 15073/3; 15091/3; 15109/3; 15127/3; 15145/3; 15163/3; 15181/3; 15199/3; 15217/3; 15235/3; 15253/3; 15271/3; 15289/3; 15307/3; 15325/3; 15343/3; 15361/3; 15379/3; 15397/3; 15415/3; 15433/3; 15451/3; 15469/3; 15487/3; 15505/3; 15523/3; 15541/3; 15559/3; 15577/3; 15595/3; 15613/3; 15631/3; 15649/3; 15667/3; 15685/3; 15703/3; 15721/3; 15739/3; 15757/3; 15775/3; 15793/3; 15811/3; 15829/3; 15847/3; 15865/3; 15883/3; 15901/3; 15919/3; 15937/3; 15955/3; 15973/3; 15991/3; 16009/3; 16027/3; 16045/3; 16063/3; 16081/3; 16099/3; 16117/3; 16135/3; 16153/3; 16171/3; 16189/3; 16207/3; 16225/3; 16243/3; 16261/3; 16279/3; 16297/3; 16315/3; 16333/3; 16351/3; 16369/3; 16387/3; 16405/3; 16423/3; 16441/3; 16459/3; 16477/3; 16495/3; 16513/3; 16531/3; 16549/3; 16567/3; 16585/3; 16603/3; 16621/3; 16639/3; 16657/3; 16675/3; 16693/3; 16711/3; 16729/3; 16747/3; 16765/3; 16783/3; 16801/3; 16819/3; 16837/3; 16855/3; 16873/3; 16891/3; 16909/3; 16927/3; 16945/3; 16963/3; 16981/3; 17000/3; 17018/3; 17036/3; 17054/3; 17072/3; 17090/3; 17108/3; 17126/3; 17144/3; 17162/3; 17180/3; 17198/3; 17216/3; 17234/3; 17252/3; 17270/3; 17288/3; 17306/3; 17324/3; 17342/3; 17360/3; 17378/3; 17396/3; 17414/3; 17432/3; 17450/3; 17468/3; 17486/3; 17504/3; 17522/3; 17540/3; 17558/3; 17576/3; 17594/3; 17612/3; 17630/3; 17648/3; 17666/3; 17684/3; 17702/3; 17720/3; 17738/3; 17756/3; 17774/3; 17792/3; 17810/3; 17828/3; 17846/3; 17864/3; 17882/3; 17900/3; 17918/3; 17936/3; 17954/3; 17972/3; 17990/3; 18008/3; 18026/3; 18044/3; 18062/3; 18080/3; 18098/3; 18116/3; 18134/3; 18152/3; 18170/3; 18188/3; 18206/3; 18224/3; 18242/3; 18260/3; 18278/3; 18296/3; 18314/3; 18332/3; 18350/3; 18368/3; 18386/3; 18404/3; 18422/3; 18440/3; 18458/3; 18476/3; 18494/3; 18512/3; 18530/3; 18548/3; 18566/3; 18584/3; 18602/3; 18620/3; 18638/3; 18656/3; 18674/3; 18692/3; 18710/3; 18728/3; 18746/3; 18764/3; 18782/3; 18800/3; 18818/3; 18836/3; 18854/3; 18872/3; 18890/3; 18908/3; 18926/3; 18944/3; 18962/3; 18980/3; 18998/3; 19016/3; 19034/3; 19052/3; 19070/3; 19088/3; 19106/3; 19124/3; 19142/3; 19160/3; 19178/3; 19196/3; 19214/3; 19232/3; 19250/3; 19268/3; 19286/3; 19304/3; 19322/3; 19340/3; 19358/3; 19376/3; 19394/3; 19412/3; 19430/3; 19448/3; 19466/3; 19484/3; 19502/3; 19520/3; 19538/3; 19556/3; 19574/3; 19592/3; 19610/3; 19628/3; 19646/3; 19664/3; 19682/3; 19700/3; 19718/3; 19736/3; 19754/3; 19772/3; 19790/3; 19808/3; 19826/3; 19844/3; 19862/3; 19880/3; 19898/3; 19916/3; 19934/3; 19952/3; 19970/3; 19988/3; 20006/3; 20024/3; 20042/3; 20060/3; 20078/3; 20096/3; 20114/3; 20132/3; 20150/3; 20168/3; 20186/3; 20204/3; 20222/3; 20240/3; 20258/3; 20276/3; 20294/3; 20312/3; 20330/3; 20348/3; 20366/3; 20384/3; 20402/3; 20420/3; 20438/3; 20456/3; 20474/3; 20492/3; 20510/3; 20528/3; 20546/3; 20564/3; 20582/3; 20600/3; 20618/3; 20636/3; 20654/3; 20672/3; 20690/3; 20708/3; 20726/3; 20744/3; 20762/3; 20780/3; 20798/3; 20816/3; 20834/3; 20852/3; 20870/3; 20888/3; 20906/3; 20924/3; 20942/3; 20960/3; 20978/3; 20996/3; 21014/3; 21032/3; 21050/3; 21068/3; 21086/3; 21104/3; 21122/3; 21140/3; 21158/3; 21176/3; 21194/3; 21212/3; 21230/3; 21248/3; 21266/3; 21284/3; 21302/3; 21320/3; 21338/3; 21356/3; 21374/3; 21392/3; 21410/3; 21428/3; 21446/3; 21464/3; 21482/3; 21500/3; 21518/3; 21536/3; 21554/3; 21572/3; 21590/3; 21608/3; 21626/3; 21644/3; 21662/3; 21680/3; 21698/3; 21716/3; 21734/3; 21752/3; 21770/3; 21788/3; 21806/3; 21824/3; 21842/3; 21860/3; 21878/3; 21896/3; 21914/3; 21932/3; 21950/3; 21968/3; 21986/3; 22004/3; 22022/3; 22040/3; 22058/3; 22076/3; 22094/3; 22112/3; 22130/3; 22148/3; 22166/3; 22184/3; 22202/3; 22220/3; 22238/3; 22256/3; 22274/3; 22292/3; 22310/3; 22328/3; 22346/3; 22364/3; 22382/3; 22400/3; 22418/3; 22436/3; 22454/3; 22472/3; 22490/3; 22508/3; 22526/3; 22544/3; 22562/3; 22580/3; 22598/3; 22616/3; 22634/3; 22652/3; 22670/3; 22688/3; 22706/3; 22724/3; 22742/3; 22760/3; 22778/3; 22796/3; 22814/3; 22832/3; 22850/3; 22868/3; 22886/3; 22904/3; 22922/3; 22940/3; 22958/3; 22976/3; 22994/3; 23012/3; 23030/3; 23048/3; 23066/3; 23084/3; 23102/3; 23120/3; 23138/3; 23156/3; 23174/3; 23192/3; 23210/3; 23228/3; 23246/3; 23264/3; 23282/3; 23300/3; 23318/3; 23336/3; 23354/3; 23372/3; 23390/3; 23408/3; 23426/3; 23444/3; 23462/3; 23480/3; 23498/3; 23516/3; 23534/3; 23552/3; 23570/3; 23588/3; 23606/3; 23624/3; 23642/3; 23660/3; 23678/3; 23696/3; 23714/3; 23732/3; 23750/3; 23768/3; 23786/3; 23804/3; 23822/3; 23840/3; 23858/3; 23876/3; 23894/3; 23912/3; 23930/3; 23948/3; 23966/3; 23984/3; 24002/3; 24020/3; 24038/3; 24056/3; 24074/3; 24092/3; 24110/3; 24128/3; 24146/3; 24164/3; 24182/3; 24200/3; 24218/3; 24236/3; 24254/3; 24272/3; 24290/3; 24308/3; 24326/3; 24344/3; 24362/3; 24380/3; 24398/3; 24416/3; 24434/3; 24452/3; 24470/3; 24488/3; 24506/3; 24524/3; 24542/3; 24560/3; 24578/3; 24596/3; 24614/3; 24632/3; 24650/3; 24668/3; 24686/3; 24704/3; 24722/3; 24740/3; 24758/3; 24776/3; 24794/3; 24812/3;

C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology  
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix  
F;37-96/Domain: von Willebrand factor type C repeat homology <VMC>  
F;1263-1492/Domain: fibrillar collagen carboxyl-terminal homology <PCC>

Query Match      45.3%; Score 488.5; DB 2; Length 1492;  
Best Local Similarity    41.5%; Pred.No. 2.5e-23;  
Matches 103; Conservative 18; Mismatches 64; Indels 63; Gaps 4;

QY                 1   GIRLKTGKEGEDGFPGFKDGMWIKGDGEIGPPRCEDEGPFGKRGPNGGDP--- 57  
Db                 | : . | : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :  
446   GPQATGLPKKETGPDGTAGTAGEHPKGEIGSAGPQGAPGAGEGRGARGFEGAA 505  
QY                 58   GLPLPPPGEKGLGVPLGYFGRQGPQRKSIGFPFGANGCKGRGTTPGKP----- 108  
Db                 ||||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :  
506   GPLPPGERGAPGNRGFPDGDLGAKGAPGARGVPLGGPKGNGSDPRPGRPGLPGAR 565  
QY                 109 -----GPRGORGBT-----GPRGERGPRGITCKPDKPKNS----- 138  
Db                 ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :  
566   GLTRPGDAGPOGKVGPSGAAGEDGRPGPPQPQCARGQQVGMPFGPKGANGPFKGAGBK 625  
QY                 139 -----GGDGPAGPGEERGPNPGPQGTGTFGPKGPRGPKDXGLP 177  
Db                 : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :  
626   GLGAPGLRLPKDGTGAGPNPAGPAGERGEQQPGPSGQLGPLPGSPSGGGKP 685  
QY                 178   GHPCQRCGE 185  
Db                 ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :

686   GDQGVPEE 693

RESULT 15  
CGHU7L  
collagen alpha 1(III) chain precursor - human  
N;Alternate names: procollagen alpha 1(III) chain  
C;Species: Homo sapiens (man)  
C;Date: 24-Apr-1984 #sequence revision 01-Sep-1995 #text change 21-Jul-2000  
C;Accession: S05272; S04642; PE0011; S01726; A04887; A90399; A94562; I51868; S59511; A90 submitted to the EMBL Data Library, February 1989  
R;Prockop, D.J.  
A;Reference number: S05272  
A;Accession: S05272  
A>Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-1240,'v',1242-1466 <PRC>  
A;Cross-references: EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058  
R;Ala-Kokko, L.; Kontusaari, S.; Baldwin, C.T.; Kuivaniemi, H.; Prockop, D.J. Biochem. J. 260, 509-516, 1989  
A>Title: Structure of cDNA clones coding for the entire prepro-alpha(III) chain of humerences.

A;Reference number: S04642; MUID:89350838; PMID:2764886  
A;Accession: S04642  
A;Molecule type: mRNA  
A;Residues: 1-1196 <ALA>  
A;Cross-references: EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058  
A>Note: the complete sequence is not shown  
R;Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F. Gene 78, 255-265, 1989  
A>Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene (COL3A1).  
A;Reference number: PE0011; MUID:89378752; PMID:2777083  
A;Accession: PE0011  
A;Molecule type: DNA  
A;Residues: 1-176 <BSN>  
A;Cross-references: GB:M26939; NID:g180813; PIDN:AAA52040.1; PID:g180814  
R;Tomon, P.D.; Ricca, G.A.; de Crombrughe, B. Nucleic Acids Res. 16, 7201, 1988  
A>Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human proalpha1(I) procollagen.  
A;Reference number: S01726; MUID:88303360; PMID:3405773  
A;Accession: S01726  
A;Molecule type: mRNA  
A;Residues: 1-170 <TON>  
A;Cross-references: EMBL:X07240; NID:g30060; PIDN:CAA30229.1;PID:g30061  
A>Note: the authors translated the codon CAG for residue 154 as His  
R;Janeczko, R.A.; Ramirez, F.



A;Molecule type: mRNA  
A;Residues: 861-1015 <COL>  
A;Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:G180878; PIDN:AAB59383.1; PID:9  
A;Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos syn  
R;Mankoo, B.S.; Dalglish, R.  
Nucleic Acids Res. 16, 2337, 1988  
A;Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.  
A;Reference number: S02119; MUID:88189827; PMID:3357782  
A;Status: translation not shown  
A;Molecule type: mRNA  
A;Residues: 950-1018, 'S', 1020-1183, 'Y', 1020-1183, 'S', 1185-1466 <MAN>  
A;Cross-references: EMBL:X06700; NID:G30053; PIDN:CAA9886.1; PID:G30054  
R;Seyer, J.M.; Kang, A.H.  
Biochemistry 20, 2621-2627, 1981  
A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from ty  
A;Reference number: A90446; MUID:81208139; PMID:7016180  
A;Accession: A90446  
A;Molecule type: protein  
A;Residues: 965-979, 'A', 981-984, 'PS', 987, 'QN', 990-1096, 'P', 1098-1152, 'AT', 1155, 'S', 1157-  
A;Experimental source: liver  
R;Loidl, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenbloom, J.; Mye  
Nucleic Acids Res. 12, 9383-9394, 1984  
A;Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollage  
A;Reference number: A93551; MUID:85087944; PMID:6096827  
A;Accession: A93551  
A;Molecule type: mRNA  
A;Residues: 1065-1155, 'P', 1157-1466 <LOI>  
A;Cross-references: EMBL:X01655; EMBL:X01742; NID:G29584; PIDN:CAA25821.1  
R;Miskulin, M.; Dalglish, R.; Kluge-Beckerman, B.; Rennard, S.E.; Tolstoshev, P.; Brant  
Biochemistry 25, 1408-1413, 1986  
A;Title: Human type III collagen gene expression is coordinately modulated with the type  
A;Reference number: I52393; MUID:86187804; PMID:3754462  
A;Accession: I52393  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1161-1200 <MS>  
A;Cross-references: GB:M13146; NID:G180415; PIDN:AAAS2003.1; PID:G180416  
R;Emanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.  
Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985  
A;Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm  
A;Reference number: I59025; MUID:85216505; PMID:3858826  
A;Accession: I79359  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1165-1196 <EMA>  
A;Cross-references: GB:M11134; NID:G180417; PIDN:AAAS2004.1; PID:G180418  
R;Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sippola, M.; Ramirez, F.  
J. Biol. Chem. 260, 4357-4363, 1985  
A;Title: Isolation of cDNA and genomic clones encoding human pro-alpha(III) collagen. F  
A;Reference number: A92516; MUID:85157600; PMID:2579949  
A;Accession: A92516  
A;Molecule type: DNA  
A;Residues: 1176-1240, 'V', 1242-1356, 'P', 1358-1466 <CHU>  
A;Cross-references: GB:M10615; GB:M10793; GB:M10794; GB:M10795; GB:M10796; GB:M10797; GB  
A;Experimental source: liver  
A;Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given d  
ation  
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently C  
C;Genetics:  
A;Gene: GDB:COL3A1  
A;Cross-references: GDB:118729; OMIM:120180  
A;Map position: 2q31-2q31  
A;Introns: 27/1, 94/3, 111/3, 149/3, 176/3, 554/3, 587/3, 1175/3, 1275/3, 1337/3, 1418/3  
A;Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Dan  
C;Complex: type III collagen is a homotrimer of monomers initially linked by disulfide b  
er of their length, is formed with desmosine cross-links made from lysine and allylsine x  
C;Function:  
A;Description: structural component of extracellular fibrous polymer that maintains inte  
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
C;Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hyd  
F;1-23/Domain: signal sequence #status predicted <SIG>

F;24-153/Domain: amino-terminal propeptide #status predicted <PRO>  
F;31-91/Domain: von Willebrand factor type C repeat homology <VWC>  
F;154-1221/Product: collagen alpha 1(III) chain #status predicted <MAT>  
F;154-167/Region: amino-terminal nonhelical telopeptide  
F;168-1196/Region: helical  
F;1091-1093/Region: cell attachment (R-G-D) motif  
F;1197-1221/Region: carboxyl-terminal nonhelical telopeptide  
F;1222-1466/Domain: carboxyl-terminal propeptide #status predicted <CPR>  
F;1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <FCC>  
F;24/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted  
F;153-154/Cleavage site: Pro-Gln (procollagen N-endopeptidase) #status predicted  
F;154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted  
F;161,1212/Modified site: allylsine (Lys) #status predicted  
F;263/Binding site: carboxylate (Lys) (covalent) #status experimental  
F;584,1094/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental  
F;948-949/Cleavage site: Gly-Ile (collagenase) #status experimental  
F;1106/Binding site: carbohydrate (Lys) (covalent) #status predicted

Query Match 45.1%; Score 486.5; DB 1; Length 1466;  
Best Local Similarity 46.4%; Pred. No. 3.2e-23;  
Matches 98; Conservative 13; Mismatches 73; Indels 27; Gaps 3;

Qy	1	GIRGLKGTGKEKEDGFPFGKDMGI	---	KGDRGETGPPCPRGEDGEGPKGKGNGDP	57
Db	711	GPPGPPCAAATPLOGMPGGERGGLGSPGPKDGEFGPGADGVPKDGSRGTGTPTGPP	770		
Qy	58	GLPLGPPGEGKLGVPGLPGVP	-----	GROGPKGSIGFGFPFGANGKGGRTGPKP	108
Db	771	GPAGQPDKGEGGAPGLGIAGPRGSPGERGTGPPGAGFPAGQNGEPGKGERGAP	830		
Qy	109	GPRGQ	-----	RGPTGPRGPRGITGKPGPKNSGGDGSPAGPPGGERGN	153
Db	831	GEKGEGLPPGVAGPPCGSGPAGPPGQVKGERGSGPGCAAGFPAGRLGPPGSGNGP	890		
Qy	154	GPOGPTGFPKPGKPPGPKDGLPHPGQR	184		
Db	891	GPPGSPGSKDGGPPGAGNTGAPSGPVS	921		

Search completed: May 3, 2004, 13:11:49  
Job time : 9.09953 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 13:01:53 ; Search time 6.07465 Seconds  
(without alignments)  
1594.339 Million cell updates/sec

Title: US-09-775-964-6

Perfect score: 1079

Sequence: 1 GIKGLKGTGKEDGDFPGF.....PPGPPGKGLPGHGORGET 186

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1079	100.0	1838	1 CA15 HUMAN	P20308 homo sapien
2	953	88.3	911	1 CA1B BOVIN	Q28083 bos taurus
3	953	88.3	1806	1 CA1B HUMAN	P12107 homo sapien
4	945	87.6	1804	1 CA1B MOUSE	Q61245 mus musculus
5	907	84.1	1736	1 CA2B HUMAN	P13942 homo sapien
6	901	83.5	1736	1 CA2B MOUSE	Q64739 mus musculus
7	737	68.3	1745	1 CA35 HUMAN	P25940 homo sapien
8	509.5	47.2	1049	1 CA13 BOVIN	P04258 bos taurus
9	501.5	46.5	1685	1 CA54 HUMAN	P29400 homo sapien
10	501	46.4	1669	1 CA14 MOUSE	P02463 mus musculus
11	496	46.0	674	1 CA14 CHICK	P08125 gallus gall
12	493.5	45.7	1262	1 CA13 CHICK	P12105 gallus gall
13	491	45.5	1496	1 CA25 HUMAN	P05997 homo sapien
14	488.5	45.3	1019	1 CA16 CHICK	P20785 gallus gall
15	487.5	45.2	1464	1 CA13 MOUSE	P08121 mus musculus
16	487	45.1	1366	1 CA21 HUMAN	P08123 homo sapien
17	486.5	45.1	1466	1 CA13 HUMAN	P02461 homo sapien
18	486	45.0	1669	1 CA14 HUMAN	P02462 homo sapien
19	485.5	45.0	674	1 CA1A BOVIN	P23206 bos taurus
20	480	44.5	1372	1 CA21 RAT	P02466 rattus norv
21	479.5	44.4	471	1 CA34 BOVIN	Q28084 bos taurus
22	479.5	44.4	671	1 CA11 RAT	P02454 rattus norv
23	479.5	44.4	1516	1 CA1H HUMAN	P39060 homo sapien
24	478.5	44.3	1464	1 CA11 HUMAN	P02452 homo sapien
25	477.5	44.3	680	1 CA1A HUMAN	Q03692 homo sapien
26	477.5	44.3	1453	1 CA11 MOUSE	P11087 mus musculus
27	477	44.2	747	1 CA12 BOVIN	P02459 bos taurus
28	477	44.2	1690	1 CA44 HUMAN	P53420 homo sapien
29	476.5	44.2	1418	1 CA12 HUMAN	P02458 homo sapien
30	476.5	44.2	1460	1 CA11 CANEA	Q9x917 canis famil
31	476	44.1	779	1 CA11 BOVIN	P02453 bos taurus
32	476	44.1	1366	1 CA21 CANFA	O46392 canis famil
33	475.5	44.1	1355	1 CA21 RANCA	O42350 rana catesb

#### ALIGNMENTS

##### RESULT 1

ID	CA15_HUMAN	STANDARD;	PRT;	1838 AA.
AC	P20908;			
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-MAY-1992 (Rel. 22, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Collagen alpha 1(V) chain precursor.			
GN	COL5A1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 556-565.			
RX	MEDLINE=91302336; PubMed=2071595;			
RA	Takahara K., Seto Y., Okasawa K., Okamoto N., Noda A., Yaoi Y.,			
RA	Kato I.;			
RT	"Complete primary structure of human collagen alpha 1 (V) chain.";			
RL	J. Biol. Chem. 266:13124-13129(1991).			
RN	[2]			
RP	SEQUENCE OF 621-822.			
RX	TISSUE=Chorioamniotic membrane;			
RX	MEDLINE=89227189; PubMed=2496661;			
RA	Seyer J.M., Kang A.H.;			
RT	"Covalent structure of collagen: amino acid sequence of three			
RT	cyanoen bromide-derived peptides from human alpha 1(V) collagen			
RT	chain.";			
RL	Arch. Biochem. Biophys. 271:120-129(1989).			
RN	[3]			
RP	SEQUENCE OF 823-950, AND HEPARIN-BINDING.			
RX	MEDLINE=90366601; PubMed=2203476;			
RA	Yaoi Y., Hashimoto K., Koitabashi H., Takahara K., Ito M., Kato I.;			
RT	"Primary structure of the heparin-binding site of type V collagen.";			
RL	Biochim. Biophys. Acta 1035:139-145(1990).			
RN	[4]			
RP	SEQUENCE OF 556-571.			
RX	TISSUE=Placenta;			
RX	MEDLINE=92239022; PubMed=1571108;			
RA	Mann K.;			
RT	"Isolation of the alpha 3-chain of human type V collagen and			
RT	characterization by partial sequencing";			
RL	Biol. Chem. Hoppe-Seyler 373:69-75(1992).			
RN	[5]			
RP	SEQUENCE OF 565-576; 756-772; 1012-1029; 1219-1232 AND 1465-1477.			
RX	TISSUE=Chorioamniotic membrane;			
RA	MEDLINE=94237164; PubMed=8181482;			
RA	Moradi-Ameli M., Rousseau J.C., Klemm J.P., Champliand M.P.,			
RA	Boutillon M.M., Bernillon J., Wallach J.M., van der Rest M.;			
RT	"Diversity in the processing events at the N-terminus of type-V			
RT	collagen.";			
RL	Eur. J. Biochem. 221:987-995(1994).			
RN	[6]			
RP	DISEASE, AND VARIANT EDS-I SER-1639.			
RX	MEDLINE=97195540; PubMed=9042913;			
RA	de Paape A., Nuytinck L., Hausser I., Anton-Lamprecht I.,			

34	475.5	44.1	1453	1	CA11 CHICK	P02457 gallus gall
35	475	44.0	744	1	CA18 HUMAN	P27658 homo sapien
36	475	44.0	1364	1	CA21 BOVIN	P02465 bos taurus
37	475	44.0	2944	1	CA17 HUMAN	Q02388 homo sapien
38	472	43.7	1459	1	CA12 MOUSE	P28481 mus musculus
39	471.5	43.7	1774	1	CA1H MOUSE	P39061 mus musculus
40	471	43.7	680	1	CA1A MOUSE	Q05306 mus musculus
41	471	43.7	744	1	CA18 RABIT	P14282 oryctolagus
42	471	43.7	1763	1	CA34 ASCSU	P27393 ascaris suu
43	469	43.5	743	1	CA18 MOUSE	Q00780 mus musculus
44	469	43.5	1028	1	CA16 HUMAN	P12109 homo sapien
45	469	43.5	1143	1	CA11 HUMAN	Q14993 homo sapien



Query Match 100.0%; Score 1079; DB 1; Length 1838;  
Best Local Similarity 100.0%; Pred. No. 4.1e-55;  
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GIRLGLKGTGKGEKEDGFGPKGDMGKIGDRGEGTGPGRGEGDPEGPKRGCGNGDGPGL 60  
DB 799 GIRLGLKGTGKGEKEDGFGPKGDMGKIGDRGEGTGPGRGEGDPEGPKRGCGNGDGPGL 858

QY 61 GPPGEGKGLGVPGLPGYRQGRQKPGKSGTGFPFGANGKGGRTGPKGPRGQGTGPR 120  
DB 859 GPPGEGKGLGVPGLPGYRQGRQKPGKSGTGFPFGANGKGGRTGPKGPRGQGTGPR 918

QY 121 GERGPRGITKPGKNSGSDGDPAGPPGPRGPNPGPGTGTGFPKPGPPGPKGDLPGHP 180  
DB 919 GERGPRGITKPGKNSGSDGDPAGPPGPRGPNPGPGTGTGFPKPGPPGPKGDLPGHP 978

QY 181 GORGET 186  
DB 979 GORGET 984

RESULT 2  
CA1B\_BOVIN STANDARD; PRT; 911 AA.  
AC Q28083;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Collagen alpha 1(XI) chain (fragment).  
GN COL1A1.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxId=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Smooth muscle;  
RX MEDLINE=92078200; PubMed=1744123;  
RA Brown K.E., Lawrence R., Sonenshein G.E.;  
RT "Concerted modulation of alpha 1(XI) and alpha 2(V) collagen mRNAs in  
RT bovine vascular smooth muscle cells."  
RL J. Biol. Chem. 266:23268-23273(1991).  
CC -!- FUNCTION: May play an important role in fibrillogenesis by  
CC controlling lateral growth of collagen II fibrils.  
CC -!- SUBUNIT: Trimers composed of three different chains: alpha 1(XI),  
CC alpha 2(XI), and alpha 3(XI). Alpha 3(XI) is a post-translational  
CC modification of alpha 1(II). Alpha 1(V) can also be found instead  
CC of alpha 3(XI)=1(II) (By similarity).  
CC -!- PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.  
CC -!- SIMILARITY: HIGH, TO ALPHA 1(V) AND ALPHA 3(V) CHAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M82977; AAA30369.1; --  
DR FTR; S18251; S18251.  
DR InterPro; IPR008161; Clg\_helix.  
DR InterPro; IPR008160; Collagen.  
DR Pfam; PF01391; Collagen; 11.  
DR ProDom; PD000007; Clg\_helix; 2.  
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Glycoprotein; Collagen.  
FT NON\_TER 1 1  
PROPEP <1 278 AMINO-TERMINAL PROPEPTIDE (POTENTIAL).

FT CHAIN 279 >911 COLLAGEN ALPHA 1(XI) CHAIN.  
FT DOMAIN <1 186 NONHELICAL REGION.  
FT DOMAIN 187 275 TRIPLE-HELICAL REGION (INTERRUPTED).  
FT DOMAIN 276 278 SHORT NONHELICAL SEGMENT.  
FT DOMAIN 279 295 TELOPEPTIDE.  
FT DOMAIN 296 >911 TRIPLE-HELICAL REGION.  
FT SITE 379 379 CROSSLINKING.  
FT NON\_TER 911 911  
SQ SEQUENCE 911 AA; 89259 MW; C05C4B3350749CFC CRC64;

Query Match 88.3%; Score 953; DB 1; Length 911;  
Best Local Similarity 88.2%; Pred. No. 3.7e-48;  
Matches 164; Conservative 4; Mismatches 18; Indels 0; Gaps 0;

QY 1 GIRLGLKGTGKGEKEDGFGPKGDMGKIGDRGEGTGPGRGEGDPEGPKRGCGNGDGPGL 60  
DB 536 GVRGLKSGKGEKEDGFGPKGDMGKIGDRGEGTGPGRGEGDPEGPKRGCGNGDGPGL 595

QY 61 GPPGEGKGLGVPGLPGYRQGRQKPGKSGTGFPFGANGKGGRTGPKGPRGQGTGPR 120  
DB 596 GQAGEKGLGVPGLPGYRQGRQKPGKSGTGFPFGANGKGGRTGPKGPRGQGTGPR 655

QY 121 GERGPRGITKPGKNSGSDGDPAGPPGPRGPNPGPGTGTGFPKPGPPGPKGDLPGHP 180  
DB 656 GSRGARGPTGKPGKSGTGSDGDPGPPGPRGPNPGPGTGTGFPKPGPPGPKGDLPGHP 715

QY 181 GORGET 186  
DB 716 GORGET 721

RESULT 3  
CA1B\_HUMAN STANDARD; PRT; 1806 AA.  
AC P12107; Q14034; Q9UIT4; Q9UIT5; Q9UIT6;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Collagen alpha 1(XI) chain precursor.  
GN COL1A1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90202924; PubMed=1690726;  
RA Yoshioka H., Ramirez F.;  
RT "Pro-alpha 1(XI) collagen. Structure of the amino-terminal propeptide  
RT and expression of the gene in tumor cell lines."  
RL J. Biol. Chem. 265:6423-6426(1990).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORMS A; B AND C), AND VARIANTS ST12/MARSHALL  
RP SYNDROME ARG-676; 921-GLN--PRO-926 DEL; 1313-PHE--GLY-1315 DEL AND  
RP VAL-1516.  
RX MEDLINE=20455728; PubMed=10486316;  
RA Annunen S., Koerkoe J., Czarny M., Warman M.L., Brunner H.G.,  
RA Kaeserlaenen H., Mulliken J.B., Tranebjaerg L., Brooks D.G.,  
RA Cox G.F., Cruysberg J.R., Curtis M.A., Davenport S.L.H.,  
RA Friedrich C.A., Kaitila I., Krawczynski M.R., Latos-Bielenska A.,  
RA Mukai S., Olsen B.R., Shinno N., Somer M., Vikkula M., Zlotogora J.,  
RA Prockop D.J., Ala-Kokko L.;  
RT "Splicing mutations of 54-bp exons in the COL1A1 gene cause Marshall  
RT syndrome, but other mutations cause overlapping Marshall/Stickler  
RT phenotypes."  
RL Am. J. Hum. Genet. 65:974-983(1999).  
RN [3]  
RP SEQUENCE OF 538-1806 FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=89034222; PubMed=3182841;  
RA Bernard M., Yoshioka H., Rodriguez E., van der Rest M., Kimura T.,  
RA Ninomiya Y., Olsen B.R., Ramirez F.;  
RT "Cloning and sequencing of pro-alpha 1 (XI) collagen cDNA  
RT demonstrates that type XI belongs to the fibrillar class of collagens

and reveals that the expression of the gene is not restricted to cartilaginous tissue.";  
J. Biol. Chem. 263:17159-17166(1988).  
[4]  
ALTERNATIVE SPLICING.  
RT TISSUE=Blood;  
RX MEDLINE=95238468; PubMed=7721876;  
RA Zhidkova N.I., Justice S.K., Mayne R.;  
RT "Alternative mRNA processing occurs in the variable region of the  
RT pro-alpha 1(XI) and pro-alpha 2(XI) collagen chains.";  
RL J. Biol. Chem. 270:9486-9493(1995).  
[5]  
VARIANT STL2 VAL-625.  
RX MEDLINE=97026296; PubMed=8872475;  
RA Richards A.J., Yates J.R.W., Williams R., Payne S.J., Pope F.M.,  
RT Scott J.D., Snead M.P.;  
RT "A family with Stickler syndrome type 2 has a mutation in the COL11A1  
RT gene resulting in the substitution of glycine 97 by valine in  
RT alpha-1(XI) collagen.";  
RL Hum. Mol. Genet. 5:1339-1343(1996).  
CC -|- FUNCTION: May play an important role in fibrillogenesis by  
CC controlling lateral growth of collagen II fibrils.  
CC -|- SUBUNIT: Trimers composed of three different chains: alpha 1(XI),  
CC alpha 2(XI), and alpha 3(XI). Alpha 3(XI) is a post-translational  
CC modification of alpha 1(II). Alpha 1(V) can also be found instead  
CC of alpha 3(XI)=1(II).  
CC -|- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=3;  
CC Comment=Additional isoforms seem to exist. There is alternative  
CC usage of exon IIA or exon IIB. Transcripts containing exon IIA  
CC or IIB are present in cartilage, but exon IIB is preferentially  
CC utilized in transcripts from tendon;  
CC Name=A;  
CC IsoId=P12107-1; Sequence=Displayed;  
CC Name=B;  
CC IsoId=P12107-2; Sequence=VSP\_001145;  
CC Name=C;  
CC IsoId=P12107-3; Sequence=VSP\_001146;  
CC -|- TISSUE SPECIFICITY: Cartilage, placenta and some tumor or virally  
CC transformed cell lines. Isoforms using exon IIA or IIB are found  
CC in the cartilage while isoforms using only exon IIB are found in  
CC the tendon.  
CC -|- PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -|- DISEASE: Defects in COL11A1 are the cause of Stickler syndrome  
CC type 2 (STL2) [MIM:604841]; also known as Stickler syndrome  
CC vitreous type 2, or beaded vitreous type, due to the presence of  
CC irregularly thickened fiber bundles throughout vitreous cavity.  
CC Stickler syndrome (hereditary progressive arthro-ophthalmopathy)  
CC is an autosomal dominant disorder characterized by progressive  
CC myopia beginning in the first decade of life, vitreo-retinal  
CC degeneration, retinal detachment, cleft palate, midfacial  
CC hypoplasia, osteoarthritis, and sensorineural hearing loss.  
CC -|- DISEASE: Defects in COL11A1 are the cause of Marshall syndrome  
CC [MIM:154780], an autosomal dominant disorder with ocular, oro-  
CC facial, auditory and skeletal manifestations. It shares several  
CC features with Stickler syndrome, such as midfacial hypoplasia,  
CC high myopia, and sensorineural-hearing deficit.  
CC -|- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.  
CC -|- SIMILARITY: HIGH, TO ALPHA 1(V) AND ALPHA 3(V) CHAINS.  
CC -|- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; J04177; AA51891.1; -;  
DR EMBL; AF101112; AAF04724.1; -;  
DR EMBL; AF101079; AAF04724.1; JOINED.

DR EMBL; AF101080; AAF04724.1; JOINED.  
DR EMBL; AF101081; AAF04724.1; JOINED.  
DR EMBL; AF101082; AAF04724.1; JOINED.  
DR EMBL; AF101083; AAF04724.1; JOINED.  
DR EMBL; AF101084; AAF04724.1; JOINED.  
DR EMBL; AF101085; AAF04724.1; JOINED.  
DR EMBL; AF101086; AAF04724.1; JOINED.  
DR EMBL; AF101087; AAF04724.1; JOINED.  
DR EMBL; AF101088; AAF04724.1; JOINED.  
DR EMBL; AF101089; AAF04724.1; JOINED.  
DR EMBL; AF101090; AAF04724.1; JOINED.  
DR EMBL; AF101091; AAF04724.1; JOINED.  
DR EMBL; AF101092; AAF04724.1; JOINED.  
DR EMBL; AF101093; AAF04724.1; JOINED.  
DR EMBL; AF101094; AAF04724.1; JOINED.  
DR EMBL; AF101095; AAF04724.1; JOINED.  
DR EMBL; AF101096; AAF04724.1; JOINED.  
DR EMBL; AF101097; AAF04724.1; JOINED.  
DR EMBL; AF101098; AAF04724.1; JOINED.  
DR EMBL; AF101099; AAF04724.1; JOINED.  
DR EMBL; AF101100; AAF04724.1; JOINED.  
DR EMBL; AF101101; AAF04724.1; JOINED.  
DR EMBL; AF101102; AAF04724.1; JOINED.  
DR EMBL; AF101103; AAF04724.1; JOINED.  
DR EMBL; AF101104; AAF04724.1; JOINED.  
DR EMBL; AF101105; AAF04724.1; JOINED.  
DR EMBL; AF101106; AAF04724.1; JOINED.  
DR EMBL; AF101107; AAF04724.1; JOINED.  
DR EMBL; AF101108; AAF04724.1; JOINED.  
DR EMBL; AF101109; AAF04724.1; JOINED.  
DR EMBL; AF101110; AAF04724.1; JOINED.  
DR EMBL; AF101111; AAF04724.1; JOINED.  
DR EMBL; AF101112; AAF04724.1; JOINED.  
DR EMBL; AF101079; AAF04726.1; -;  
DR EMBL; AF101080; AAF04726.1; JOINED.  
DR EMBL; AF101081; AAF04726.1; JOINED.  
DR EMBL; AF101082; AAF04726.1; JOINED.  
DR EMBL; AF101083; AAF04726.1; JOINED.  
DR EMBL; AF101084; AAF04726.1; JOINED.



Db 767 GVRGLKSGKEGEGDPGPKGDMGLKXDRGEVCQVPRGDDGPEGPKGRAGTGDGPPS 826

Qy 61 GPPGEKGLGVPLPGVPGQPKGSIGFPFGPANGKEKGGRTGPKGPRGQGTGPR 120

Db 827 QGAGEKGLGVPLPGVPGQPKGSTGFPFGPANGKEKGGRTGPKGPRGQGTGPR 886

Qy 121 GERPGRTGKPGKGNSSGDPAGPGRGPNQPGQPTGPPGKPPGPKDGLPGHP 180

Db 887 GSRGARPTGKPGPKGTSGGDPGPPGPRGQPGQPVGFPKPPGPKDGLPGHP 946

Qy 181 GQGET 186

Db 947 GQGET 952

RESULT 5

CA2B HUMAN

ID CA2B HUMAN STANDARD; PRT: 1736 AA.

DT P13942; Q07751; Q13271; Q13272; Q13273; Q39866; Q9UIP9;

DT 01-JAN-1990 (Rel. 13, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Collagen alpha 2(XI) chain precursor.

GN COL11A2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96032717; PubMed=7559422;

RA Vuorio M.M., Pihlajamaa T., Vandenbergh P., Prockop D.J.,

RA Ala-Kokko L.;

RT "The human COL11A2 gene structure indicates that the gene has not

RT evolved with the genes for the major fibrillar collagens.;"

RJ J. Biol. Chem. 270:22873-22881(1995).

RN [2]

RP SEQUENCE FROM N.A.

RA Tubby B.;

RN Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 59-807 FROM N.A.

RC TISSUE=Cartilage;

RX MEDLINE=93314796; PubMed=8325374;

RA Zhidkova N.I., Brewton R.G., Mayne R.;

RT "Molecular cloning of PARP (proline/arginine-rich protein) from human

RT cartilage and subsequent demonstration that PARP is a fragment of the

RT NH2-terminal domain of the collagen alpha 2(XI) chain.;"

RL FEBS Lett. 326:25-28(1993).

RN [4]

RP SEQUENCE OF 730-1690 FROM N.A.

RX MEDLINE=89340485; PubMed=2760050;

RA Kimura T., Cheah K.S.E., Chan S.D.H., Lui V.C.H., Mattei M.-G.,

RA van der Rest M., Ono K., Solomon E., Ninomiya Y., Olsen B.R.;

RT "The human alpha 2(XI) collagen (COL11A2) chain. Molecular cloning of

RT cDNA and genomic DNA reveals characteristics of a fibrillar collagen

RT with differences in genomic organization.;"

RL J. Biol. Chem. 264:13910-13916(1989).

RN [5]

RP SEQUENCE OF 1-537 FROM N.A.

RX MEDLINE=96435918; PubMed=8838804;

RA Lui V.C., Ng L.J., Sat E.W., Cheah K.S.;

RT "The human alpha 2(XI) collagen gene (COL11A2): completion of coding

RT information, identification of the promoter sequence, and precise

RT localization within the major histocompatibility complex reveal

RT overlap with the KES5 gene.;"

RL Genomics 32:401-412(1996).

RN [6]

RP ALTERNATIVE SPLICING.

RX MEDLINE=95238468; PubMed=7721876;

RA Zhidkova N.I., Justice S.K., Mayne R.;

RT "Alternative mRNA processing occurs in the variable region of the

RT pro-alpha 1(XI) and pro-alpha 2(XI) collagen chains.;"

RL

RL J. Biol. Chem. 270:9486-9493(1995).

RN [7]

RP DISEASE.

RX MEDLINE=20143361; PubMed=10677296;

RA Melkonian M., Brunner H.G., Manouvrier S., Hennekam R.,

RA Superti-Furga A., Kaeerlaeinen H., Pauli R.M., van Esen T.,

RA Warman M.L., Bonaventura J., Miny P., Ala-Kokko L.;

RT "Autosomal recessive disorder otospondylocapsular dysplasia is

RT associated with loss-of-function mutations in the COL11A2 gene.;"

RL Am. J. Hum. Genet. 66:368-377(2000).

RN [8]

RP REVIEW ON VARIANTS.

RX MEDLINE=97255959; PubMed=9101290;

RA Kuivaniemi H., Tromp G., Prockop D.J.;

RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-

RT associated collagen (type IX), and network-forming collagen (type X)

RT cause a spectrum of diseases of bone, cartilage, and blood vessels.;"

RL Hum. Mutat. 9:300-315(1997).

RN [9]

RP VARIANT OSMED ARG-661.

RX MEDLINE=95163096; PubMed=7859284;

RA Vikkula M., Mariman E.C.M., Lui V.C.H., Zhidkova N.I., Tiller G.E.,

RA Goldring M.B., van Beersum S.E.C., de Waal Malefijt M.C.,

RA van den Hoogen F.H.J., Kopers H.-H., Mayne R., Cheah K.S.E.,

RA Olsen B.R., Warman M.L., Brunner H.G.;

RT "Autosomal dominant and recessive osteochondrodysplasias associated

RT with the COL11A2 locus.;"

RL Cell 80:431-437(1995).

RN [10]

RP VARIANTS GLY-593; LYS-924; LEU-879; THR-1316 AND GLN-1600.

RX MEDLINE=98254467; PubMed=9585596;

RA Koga H., Sakou T., Taketomi E., Hayashi K., Numasawa T., Harata S.,

RA Yone K., Matsunaga S., Otterud B., Inoue I., Leppert M.;

RT "Genetic mapping of ossification of the posterior longitudinal

RT ligament of the spine.;"

RL Am. J. Hum. Genet. 62:1460-1467(1998).

RN [11]

RP VARIANT WZS GLU-955.

RX MEDLINE=99021942; PubMed=9805126;

RA Pihlajamaa T., Prockop D.J., Faber J., Winterpacht A., Zabel B.,

RA Gledion A., Wiesbauer P., Spranger J., Ala-Kokko L.;

RT "Heterozygous glycine substitution in the COL11A2 gene in the original

RT patient with the Weissensbacher-Zweymueller syndrome demonstrates its

RT identity with heterozygous OSMED (nonocular Stickler syndrome).;"

RL Am. J. Med. Genet. 80:115-120(1998).

RN [12]

RP VARIANT STL3 940-GLY--PRO-948 DEL.

RX MEDLINE=98165506; PubMed=9506662;

RA Sirko-Osada D.A., Murray M.A., Scott J.A., Lavery M.A., Warman M.L.,

RA Robin N.H.;

RT "Stickler syndrome without eye involvement is caused by mutations in

RT COL11A2, the gene encoding the alpha-2(XI) chain of type XI

RT collagen.;"

RL J. Pediatr. 132:368-371(1998).

RN [13]

RP VARIANTS DFNA13 GLU-808 AND CYS-1034, AND REVISIONS TO 1031-1032.

RX MEDLINE=20047768; PubMed=10581026;

RA McGuire W.T., Prasad S.D., Griffith A.J., Kunst H.P.M., Green G.E.,

RA Shargel K.B., Runge C., Huybrechts C., Mueller R.F., Lynch E.,

RA King M.-C., Brunner H.G., Cremers C.W.R.J., Takasumi M., Li S.-W.,

RA Arita M., Mayne R., Prockop D.J., Van Camp G., Smith R.J.H.;

RT "Mutations in COL11A2 cause non-syndromic hearing loss (DFNA13).;"

RL Nat. Genet. 23:413-419(1999).

CC -!- FUNCTION: May play an important role in fibrillogenesis by

CC controlling lateral growth of collagen II fibrils.

CC -!- SUBUNIT: Trimers composed of three different chains: alpha 1(XI),

CC alpha 2(XI), and alpha 3(XI). Alpha 3(XI) is a post-translational

CC modification of alpha 1(XI). Alpha 1(XI) can also be found instead

CC of alpha 3(XI)=1(XI).

CC -!- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=8;

CC Comment=Isoforms lack exons 6, 7 or 8 or a combination of these

CC exons. Experimental confirmation may be lacking for some



CC isoforms;  
 CC Name=1; IsoId=P13942-1; Sequence=Displayed;  
 CC Name=2; IsoId=P13942-2; Sequence=VSP\_001167;  
 CC Name=3; IsoId=P13942-3; Sequence=VSP\_001168;  
 CC Name=4; IsoId=P13942-4; Sequence=VSP\_001169;  
 CC Name=5; IsoId=P13942-5; Sequence=VSP\_001167, VSP\_001168;  
 CC Name=6; IsoId=P13942-6; Sequence=VSP\_001167, VSP\_001169;  
 CC Name=7; IsoId=P13942-7; Sequence=VSP\_001168, VSP\_001169;  
 CC Name=8; IsoId=P13942-8; Sequence=VSP\_001167, VSP\_001168;  
 CC -1- PTM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.  
 CC -1- PTM: A disulfide-bonded peptide called proline/arginine-rich protein or PARP is released from the amino terminus during extracellular processing and is subsequently retained in the cartilage matrix from which it can be isolated in significant amounts.  
 CC -1- DISEASE: Defects in COL11A2 are the cause of Stickler syndrome type 3 (STL3) [MIM:184840]. STL3 is an autosomal dominant disorder characterized by oro-facial, auditory and skeletal manifestations, such as midfacial hypoplasia, cleft palate, osteoarthritis, and sensorineural hearing loss. Differently from Stickler syndrome type 1 and 2, no ocular involvement is observed. This disorder is also referred to as Stickler-like syndrome or non-ocular Stickler syndrome.  
 CC -1- DISEASE: Defects in COL11A2 are the cause of autosomal recessive otospondylocheilodysplasia (OSMED) [MIM:215150], a skeletal dysplasia accompanied by severe hearing loss. The phenotype overlaps that of autosomal dominant skeletal disorders (Stickler and Marshall syndromes) but can be distinguished by disproportionately short limbs and lack of ocular involvement.  
 CC -1- DISEASE: Defects in COL11A2 are the cause of Weissenbacher-Zweymueller syndrome (WZS) [MIM:277610], an autosomal dominant disorder allelic with STL3 and OSMED. WZS is also referred to as heterozygous OSMED.  
 CC -1- DISEASE: Defects in COL11A2 are the cause of autosomal dominant nonsyndromic sensorineural deafness type 13 (DFNA13) [MIM:601868]. Affected individuals experience progressive hearing loss beginning in the second to fourth decades, eventually making use of amplification mandatory.  
 CC -1- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.  
 CC -1- DATABASE: Contains 1 TSP N-terminal (TSPN) domain.  
 CC WWW="http://www.uit.ac.be/dnalab/hhh/".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U32169; AAC50214.1; -  
 CC EMBL; U32169; AAC50213.1; -  
 CC EMBL; U32169; AAC50215.1; -  
 CC EMBL; AL031228; CAA20240.1; -  
 CC EMBL; L18987; AAA35498.1; -  
 CC EMBL; J04974; AAA2034.1; -  
 CC EMBL; U41069; AAC17464.1; -  
 CC EMBL; U41065; AAC17464.1; JOINED.  
 CC EMBL; U41066; AAC17464.1; JOINED.  
 CC EMBL; U41067; AAC17464.1; JOINED.  
 CC Genew; HGNC:2187; COL11A2.

Query Match

84.1%; Score 907; DB 1; Length 1736;

Best Local Similarity 85.4%; Pred. No. 2.6e-45;  
 Matches 158; Conservative 7; Mismatches 20; Indels 0; Gaps 0;  
 QY 1 GIRGLKGTGKGEKGDGPGFKGDMGIKGRGEIGPPGPRGEDGEGPKGRGGNGDPGL 60  
 DB 727 GIRGLKGHKGEKGDGPGFKGDIKVGDRGEVGVPSRGEDGEGPKGRGTGTGDPGP 786  
 QY 61 GPPGEXKGLGVPLPGYPOKQPGKSGIGPPGPGANGKGGRTTPKPPRGORGPTGPR 120  
 DB 787 GLMGEXKGLGVPLPGYPOKQPGKSLGFGPGFPGASGKRGARLSGSKSPRGSRGPTGPR 846  
 QY 121 GERGPRGITKCPKCNKSGDGPAGPPGPRGPNPGOGTGFPGPKPGPPGKGLPGHP 180  
 DB 847 GQGRPGATGKSGAKTSGDGHPPGPRGUPGQPGNGFGPGKPGPPGPKDGLPGHP 906  
 QY 181 GORGE 185  
 DB 907 GORGE 911  
 RESULT 6  
 CA2B MOUSE  
 ID CA2B MOUSE STANDARD; PRT; 1736 AA.  
 AC Q64739; O61432; Q521W0;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Collagen alpha 2(XI) chain precursor.  
 GN COL11A2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 7).  
 RC STRAIN=129/SvJ.  
 RC Rowen L., Qin S., Madan A., Loretz C., James R., Dors M., Mix L.,  
 RA Hall J., Lasky S., Hood L.;  
 RA "Sequence of the mouse major histocompatibility locus class II  
 RT region.";  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 1-1678 FROM N.A. (ISOFORM 7).  
 RC STRAIN=FVB/N, and 129/Sv; TISSUE=Cartilage;  
 RX MEDLINE=97133755; PubMed=8981332;  
 RA Vandenbergh P., Vuorio M.M., Ala-Kokko L., Prockop D.J.;  
 RT "The mouse colla2 gene. Some transcripts from the adjacent rrx-beta  
 RT gene extend into the colla2 gene.";  
 RL Matrix Biol. 15:359-367(1996).  
 RN [3]  
 RP SEQUENCE OF 1-624 FROM N.A. (ISOFORMS 1; 2; 3; 4; 5 AND 6).  
 RC STRAIN=129/Sv;  
 RX MEDLINE=95138212; PubMed=7836472;  
 RA Tsunaki N., Kimura T.;  
 RT "Differential expression of an acidic domain in the amino-terminal  
 RT propeptide of mouse pro-alpha2(XI) collagen by complex alternative  
 RT splicing.";  
 RL J. Biol. Chem. 270:2372-2378(1995).  
 RN [4]  
 RP SEQUENCE OF 1-8 FROM N.A.  
 RC STRAIN=129/Sv; TISSUE=Liver;  
 RX MEDLINE=96427460; PubMed=8830784;  
 RA Tsunaki N., Kimura T., Matsui Y., Ochi T.;  
 RT "Separable cis-regulatory elements that contribute to tissue- and  
 RT site-specific alpha 2(XI) collagen gene expression in the embryonic  
 RT mouse cartilage.";  
 RL J. Cell Biol. 134:1573-1582(1996).  
 CC -1- FUNCTION: May play an important role in fibrillogenesis by  
 CC controlling lateral growth of collagen II fibrils (By similarity).  
 CC -1- SUBUNIT: Trimers composed of three different chains: alpha 1(XI),  
 CC alpha 2(XI), and alpha 3(XI). Alpha 3(XI) is a post-translational  
 CC modification of alpha 1(ii). Alpha 1(v) can also be found instead  
 CC of alpha 3(XI)=1(ii) (By similarity).

```
CC -|- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=7;
CC Comment=Additional isoforms seem to exist;
CC Name=1; Synonyms=E56789;
CC IsoId=Q64739-2; Sequence=Displayed;
CC Name=2; Synonyms=E5689;
CC IsoId=Q64739-3; Sequence=VSP_007346;
CC Name=3; Synonyms=E5789;
CC IsoId=Q64739-4; Sequence=VSP_007345;
CC Name=4; Synonyms=E569;
CC IsoId=Q64739-5; Sequence=VSP_007346; VSP_007347;
CC Name=5; Synonyms=E589;
CC IsoId=Q64739-6; Sequence=VSP_007345; VSP_007346;
CC Name=6; Synonyms=E59;
CC IsoId=Q64739-7; Sequence=VSP_007345; VSP_007346; VSP_007347;
CC Name=7;
CC IsoId=Q64739-1; Sequence=VSP_007345; VSP_007347;
CC -|- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -|- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.
CC -|- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF100956; AAC69905.1; -
CC EMBL; U16789; AAA67751.1; -
CC EMBL; U16790; AAA67752.1; -
CC EMBL; D38412; BAA18910.1; -
CC EMBL; D84066; BAA12208.1; -
CC MGI; 88447; Coll1a2.
CC InterPro; IPR008161; Clg_helix.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR008985; ConA_like_lec_gl.
CC InterPro; IPR000885; Fib_Collagen_C.
CC InterPro; IPR001791; LamTnIn_G.
CC InterPro; IPR003129; TSPN.
CC Pfam; PF01410; COLFI; 1.
CC Pfam; PF01391; Collagen; 18.
CC Pfam; PF02210; TSPN; 1.
CC ProDom; PD000007; Clg_helix; 1.
CC ProDom; PD002078; Fib_collagen_C; 1.
CC SMART; SM00038; COLFI; 1.
CC SMART; SM00282; LamG; 1.
CC SMART; SM00210; TSPN; 1.
CC KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC Glycoprotein; Collagen; Alternative splicing; Signal.
CC FT SIGNAL 1 22 POTENTIAL.
CC CHAIN 23 1500 COLLAGEN ALPHA 2 (X1) CHAIN.
CC PROPEP 1501 1736 CARBOXYL-TERMINAL PROPEPTIDE.
CC DOMAIN 31 214 TSP N-TERMINAL.
CC DOMAIN 215 486 NONHELICAL REGION.
CC DOMAIN 487 1500 TRIPLE-HELICAL REGION.
CC DOMAIN 1501 1736 NONHELICAL REGION.
CC CARBOHYD 1604 1604 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC VARSPLIC 267 292 Missing (in isoform 3, isoform 5, isoform
CC 6 and isoform 7).
CC FT FT /FTId=VSP_007345.
CC FT FT Missing (in isoform 2, isoform 4, isoform
CC 5 and isoform 6).
CC FT FT /FTId=VSP_007346.
CC FT FT Missing (in isoform 4, isoform 6 and
CC isoform 7).
CC FT FT /FTId=VSP_007347.
CC FT FT R -> L (IN REF. 2).
CC FT FT P -> S (IN REF. 3).
CC FT FT NQ -> KP (IN REF. 2).
CC FT FT V -> A (IN REF. 2).
CC FT FT

CC FT CONFLICT 536 536
CC FT CONFLICT 621 621
CC FT CONFLICT 704 705
CC FT CONFLICT 797 797

CC FT CONFLICT 843 845 TGP -> HGS (IN REF. 2).
CC FT CONFLICT 854 854 A -> S (IN REF. 2).
CC FT CONFLICT 876 876 R -> G (IN REF. 2).
CC FT CONFLICT 889 889 G -> V (IN REF. 2).
CC FT CONFLICT 922 922 G -> D (IN REF. 2).
CC FT CONFLICT 1005 1005 E -> V (IN REF. 2).
CC FT CONFLICT 1253 1253 P -> S (IN REF. 2).
CC FT CONFLICT 1386 1386 A -> T (IN REF. 2).
CC FT CONFLICT 1522 1522 I -> M (IN REF. 2).
CC SQ SEQUENCE 1736 AA; 171535 MW; 18D792D4A3387C61 CRC64;

Query Match 83.5%; Score 901; DB 1; Length 1736;
Best Local Similarity 84.9%; Pred. No. 5.7e-45;
Matches 157; Conservative 7; Mismatches 21; Indels 0; Gaps 0;

QY 1 GIRGLKTKYGEKGEDGPFKGMKIGDRBHGIPGPRGEGDGGPGKRGPGNPDGGL 60
DB 727 GIRGLKHGKGEKGEDGPFKGMKIGDRBHGIPGPRGEGDGGPGKRGPGNPDGGL 786
QY 61 GPPGKGLGVPLPGYPRGQPKSGIGPFGPGANGKGGRTGPKGPRGQRTGPR 120
DB 787 GLMGKGLGVPLPGYPRGQPKSGIGPFGPGANGKGGRTGPKGPRGQRTGPR 846
QY 121 GERGPRGITGPKPKNGSGDGPAGPGERGPNQGPTGPPGKPGPPGKGLPGHP 180
DB 847 GQRGPRGATGSKAGTSGDGPHPGGERGLPGQPGNPGFPKPGPPGKGLPGHP 906
QY 181 GQGE 185
DB 907 GQGE 911

RESULT 7
CC35_HUMAN
ID CA35_HUMAN STANDARD; PRT; 1745 AA.
AC P25940; Q9NZQ6;
DC 01-MAY-1992 (Rel. 22, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 3(V) chain precursor.
COL5A3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP SEQUENCE FROM N.A.
RC TISSUE=Heart, and Placenta;
RX MEDLINE=20187594; PubMed=10722718;
RA Imamura Y., Scott I.C., Greenspan D.S.;
RT "The pro-alpha3 (V) collagen chain. Complete primary structure,
RT expression domains in adult and developing tissues, and comparison to
RT the structures and expression domains of the other types V and XI
RT procollagen chains.";
RL J. Biol. Chem. 275:8749-8759(2000).
RN [2]
RP PRELIMINARY SEQUENCE OF 479-564; 665-709; 723-758; 787-816; 922-1008;
RP 1054-1088; 1248-1287 AND 1313-1334.
RC TISSUE=Placenta;
RX MEDLINE=92239022; PubMed=1571108;
RA Mann K.;
RT "Isolation of the alpha 3-chain of human type V collagen and
RT characterization by partial sequencing.";
RL Biol. Chem. Hoppe-Seyler 373:69-75(1992).
CC -|- FUNCTION: Type V collagen is a member of group I collagen
CC (fibrillar forming collagen). It is a minor connective tissue
CC component of nearly ubiquitous distribution. Type V collagen binds
CC to DNA, heparan sulfate, thrombospondin, heparin, and insulin.
CC -|- SUBUNIT: Trimers of two alpha 1(V) and one alpha 2(V) chains in
CC most tissues and trimers of one alpha 1(V), one alpha 2(V), and
CC one alpha 3(V) chains in placenta.
CC -|- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
```



```
FT CARBOHYD 107 107 O-LINKED (GAL. . .).
FT CARBOHYD 950 950 O-LINKED (GAL. . .).
FT DISULFID 1040 1040 INTERCHAIN.
FT DISULFID 1041 1041 INTERCHAIN.
SQ SEQUENCE 1049 AA, 93651 MW, 8EBC3D1C66EC9A3 CRC64;

Query Match 47.2%; Score 509.5; DB 1; Length 1049;
Best local similarity 38.5%; Pred. No. 8.8e-23;
Matches 109; Conservative 14; Mismatches 61; Indels 99; Gaps 6;

QY 1 GIRLXGT-----KGEKGEDGFFGKDMGKIDRGEI---GPPGPRGED 42
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
483 GLQLPSTGPPGNGKPGEPGPGKGEAGAPGIPGGKDGSCAPGERGPPGAGGPPGRGA 542
QY 43 GPEGPKG-----RGGN----- 54
Db |||:|||||:|||||:|||||:|||||:|||||:
543 GPPGEGGKAAGPPGPPGSGAGTGLQGMPERGPGPGPKDGKGEFGSSGVDGAPGKD 602
QY 55 -----GDPGLGPPGEGKGLGVPLGYP-----GRQPKGSIFFPPGPPGAN 96
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:
603 GPRGPTGPIGPPGAPGQDGKESGAPGVFGIAGPRGGPGERGEQGPPGPFAGFPAGQN 662
QY 97 GEKGGRTGKCPGRQ-----RGPTGPRGERGPRGITKPGPKGNSGD 141
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:
663 GEPGKGERGAPGEKGEKGGPPGAAGPAGSGPAGPPGQGVKGERSGPGGAAGFPGR 722
QY 142 GPAGPPGERGPPGPGTGPFGPKGPPGPKDGLPHGPGQ 184
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:
723 GPPGPGSNGNPPGSSGAPGKDGPPGPGSNGAPGSGISG 765

RESULT 9
CAS4 HUMAN STANDARD; PRT; 1685 AA.
ID -CA54 HUMAN
AC P29470: Q16006; Q16126;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 5(IV) chain precursor.
GN COL4A5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94165049; PubMed=8120014;
RA Zhou J., Leinonen A., Tryggvason K.;
RT "Structure of the human type IV collagen COL4A5 gene.";
RL J. Biol. Chem. 269:6608-6614(1994).
RN [2]
RP SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CYS-521.
RC TISSUE=Kidney;
RX MEDLINE=92316923; PubMed=1352287;
RA Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;
RT "Complete amino acid sequence of the human alpha 5 (IV) collagen chain and identification of a single-base mutation in exon 23 converting glycine 521 in the collagenous domain to cysteine in an Alport syndrome patient.";
RL J. Biol. Chem. 267:12475-12481(1992).
RN [3]
RP SEQUENCE OF 85-1685 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=90337990; PubMed=2380186;
RA Pihlajaniemi T., Pohjolainen E.R., Myers J.C.;
RT "Complete primary structure of the triple-helical region and the carboxyl-terminal domain of a new type IV collagen chain, alpha 5(IV).";
RL J. Biol. Chem. 265:13758-13766(1990).
RN [4]
RP SEQUENCE OF 924-1685 FROM N.A.
RX MEDLINE=91169491; PubMed=2004755;
RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;
```

```
RT "Characterization of the 3' half of the human type IV collagen alpha 5 gene that is affected in the Alport syndrome.";
RL Genomics 9:1-9(1991).
RN [5]
RP SEQUENCE OF 914-1685 FROM N.A.
RX MEDLINE=90160375; PubMed=1689491;
RA Hostikka S.L., Eddy R.L., Byers M.G., Hoeslytyae M., Shows T.B., Tryggvason K.;
RT "Identification of a distinct type IV collagen alpha chain with restricted kidney distribution and assignment of its gene to the locus of X chromosome-linked Alport syndrome.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).
RN [6]
RP SEQUENCE OF 1442-1471 FROM N.A.
RX MEDLINE=90252791; PubMed=2339699;
RA Myers J.C., Jones T.A., Pohjolainen E.R., Kadri A.S., Goddard A.D., Sher D., Solomon E., Pihlajaniemi T.;
RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene to the region of the X chromosome containing the Alport syndrome locus.";
RL Am. J. Hum. Genet. 46:1024-1033(1990).
RN [7]
RP SEQUENCE OF 1-20 FROM N.A.
RA Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J., Marynen P.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE OF 1258-1270 FROM N.A. (ISOFORM 2).
RX MEDLINE=94133540; PubMed=8301933;
RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H., Cassiman J.-J., Marynen P.;
RT "Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex mutation in the COL4A5 gene of an Alport patient deletes the Nc1 domain.";
RL Kidney Int. 44:1316-1321(1993).
RN [9]
RP REVIEW ON VARIANTS.
RX MEDLINE=97338662; PubMed=9195222;
RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
RT "The clinical spectrum of type IV collagen mutations.";
RL Hum. Mutat. 9:477-499(1997).
RN [10]
RP VARIANT AS SER-1564.
RX MEDLINE=91169492; PubMed=1672282;
RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L., Tryggvason K.;
RT "Single base mutation in alpha 5(IV) collagen chain gene converting a conserved cysteine to serine in Alport syndrome.";
RL Genomics 9:10-18(1991).
RN [11]
RP VARIANT AS ARG-325.
RX MEDLINE=92303559; PubMed=1376965;
RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P., Tryggvason K., Gubler M.-C., Antignac C.;
RT "Substitution of arginine for glycine 325 in the collagen alpha 5 (IV) chain associated with X-linked Alport syndrome: characterization of the mutation by direct sequencing of PCR-amplified lymphoblast cDNA fragments.";
RL Am. J. Hum. Genet. 51:135-142(1992).
RN [12]
RP VARIANT AS GLU-325.
RX MEDLINE=93244772; PubMed=1363780;
RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L., Rizzoni G.F., de Marchi M.;
RT "De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in Alport syndrome.";
RL Hum. Mol. Genet. 1:127-129(1992).
RN [13]
RP VARIANTS AS THR-1517; SER-1538 AND GLN-1563.
RX MEDLINE=94010948; PubMed=8406498;
RA Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J., Tryggvason K., Haggema-Schouten W.A.G., Roodvoets A.P., Rascher W., van Oost B.A., Smeets H.J.N.;
```

RT "Identification of four novel mutations in the COL4A5 gene of  
 RT patients with Alport syndrome.";  
 RL Genomics 17:485-489(1993).  
 RN [14]  
 RP VARIANTS AS GLU-400; VAL-406; VAL-638; ARG-653; ARG-796;  
 RP ARG-869; ARG-872 AND CYS-1241.  
 RX MEDLINE=95322976; PubMed=7599631;  
 RA Boye E., Flinter F., Zhou J., Tryggvason K., Bobrow M., Harris A.;  
 RA "Detection of 12 novel mutations in the collagenous domain of the  
 RT COL4A5 gene in Alport syndrome patients.";  
 RL Hum. Mutat. 5:197-204(1995).  
 RN [15]  
 RP VARIANT AS ARG-1649.  
 RX MEDLINE=96213750; PubMed=8651292;  
 RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,  
 RA Denison J.C., Fain P.R., Gregory M.C.;  
 RT "A mutation causing Alport syndrome with tardive hearing loss is  
 RT common in the western United States.";  
 RL Am. J. Hum. Genet. 58:1157-1165(1996).  
 RN [16]  
 RP VARIANTS AS.  
 RX MEDLINE=96213754; PubMed=8651296;  
 RA Renieri A., Bruttini M., Galli L., Zanelli P., Rossetti S.,  
 RA Turco A.E., Heiskari N., Zhou J., Gusmano R., Maesella L., Banfi G.,  
 RA Scolari F., Sesca A., Rizzoni G.F., Tryggvason K., Pignatti P.F.,  
 RA Savi M., Ballabio A., de Marchi M.;  
 RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51  
 RT exons of the COL4A5 gene.";  
 RL Am. J. Hum. Genet. 58:1192-1204(1996).  
 RN [17]  
 RP VARIANTS AS, AND VARIANTS ASP-430; SER-444; SER-619; ASN-664 AND  
 RP MET-1428.  
 RX MEDLINE=97094179; PubMed=8940267;  
 RA Knebelmann B., Breillat C., Forestier L., Arrondel C., Jacassier D.,  
 RA Glatas I., Drouot L., Deschenes G., Gruenfeld J.-P., Broyer M.,  
 RA Gubler M.-C., Antignac C.;  
 RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport  
 RT syndrome.";  
 RL Am. J. Hum. Genet. 59:1221-1232(1996).  
 RN [18]  
 RP VARIANT AS ASP-1498.  
 RX MEDLINE=96233932; PubMed=8829632;  
 RA Tverskaya S., Bobrykhina V., Tsalykova F., Ignatova M.,  
 RA Krasnopolskaya X., Evgrafov O.;  
 RT "Substitution of A1498D in noncollagen domain of  $\alpha 5(\text{IV})$  collagen  
 RT chain associated with adult-onset X-linked Alport syndrome.";  
 RL Hum. Mutat. 7:149-150(1996).  
 RN [19]  
 RP VARIANT AS GLN-1677.  
 RX MEDLINE=97295089; PubMed=9150741;  
 RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;  
 RT "Common ancestry of three Ashkenazi-American families with Alport  
 RT syndrome and COL4A5 R1677Q.";  
 RL Hum. Genet. 99:681-684(1997).  
 RN [20]  
 RP VARIANTS AS ARG-174; ARG-177; ARG-325; CYS-1410; TRP-1421; THR-1517  
 RP AND ASP-1596.  
 RX MEDLINE=98112435; PubMed=9452056;  
 RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,  
 RA Pignatti G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,  
 RA Trivelli A., Pinciaroli A.R., Ragalolo M., Rizzoni G.F., de Marchi M.;  
 RT "Missense mutations in the COL4A5 gene in patients with X-linked  
 RT Alport syndrome.";  
 RL Hum. Mutat. Suppl. 1:S106-S109(1998).  
 RN [21]  
 RP VARIANTS AS VAL-420; 456-PRO--PRO-458 DEL; ASP-573; ASP-624; ASP-635;  
 RP 802-GLY--PRO-807 DEL; ARG-869; CYS-941; SER-1030; SER-1066; ASP-1143;  
 RP ARG-1196; GLU-1261; SER-1357 AND ARG-1649.  
 RX MEDLINE=99063529; PubMed=9848783;  
 RA Martin P., Heiskari N., Zhou J., Leinonen A., Tumelius T., Hertz J.M.,  
 RA Barker D.F., Gregory M.C., Atkin C.L., Stykarsdottir U., Neumann H.,  
 RA Springate J., Shows T.B., Petersson E., Tryggvason K.;  
 RT "High mutation detection rate in the COL4A5 collagen gene in suspected

RT Alport syndrome using PCR and direct DNA sequencing.";  
 RL J. Am. Soc. Nephrol. 9:2291-2301(1998).  
 RN [22]  
 RP VARIANTS AS GLU-579; LYS-633; ASP-947; VAL-953; ARG-1107; ARG-1158;  
 RP SER-1170 AND TRP-1678, AND VARIANTS SER-444 AND ALA-739.  
 RX MEDLINE=20030197; PubMed=10561141;  
 RA Inoue Y., Nishio H., Shirakawa T., Nakanishi K., Nakamura H.,  
 RA Sumino K., Nishiyama K., Iijima K., Yoshikawa N.;  
 RT "Detection of mutations in the COL4A5 gene in over 90% of male  
 RT patients with X-linked Alport's syndrome by RT-PCR and direct  
 RT sequencing.";  
 RL Am. J. Kidney Dis. 34:854-862(1999).  
 RN [23]  
 RP VARIANT AS ARG-822.  
 Query Match 46.5%; Score 501.5; DB 1; Length 1685;  
 Best Local Similarity 47.2%; Pred. No. 3.7e-22;  
 Matches 108; Conservative 13; Mismatches 63; Indels 45; Gaps 7;  
 QY 1 GTRGLKGTGKGEQDFPGKGMGKIDGRGE-----TGPQPRGEDGP--EGPKG 49  
 DB 1193 GPPGLFGLSGQKGDGLPGIPGNPGLPGKGEFHFPGVQVQPPGSPGPALEGPKG 1252  
 QY 50 RCGPNGD-----PGPLGPP-----GEKGLGVPLPGVP---GROGPKSGIFPG 91  
 DB 1253 NFGPQGPGRPGPLPGEGPPGLPGNGIGKEKNPQGLPGLGKDGQGPGLQGNPG 1312  
 QY 92 PFGANGKEGGRGTPGKPGPRGQRGPTGPRGERGPR---GITGKPGP-----K 135  
 DB 1313 RPLGLNMGKGDPLGPGVPGFPGMKPSGVPGSAGPEGELIGPPGPPGLPGSPGQSIIK 1372  
 QY 136 GNSGDDGPPAGPPGERGPNQPGPTGPPGKGGPPGPKDGLPCHPQQR 184  
 DB 1373 GDAGPPGIPQPLGKGLPGPQGLPGTGPDPGDPGRNGLPGLFDGAGG 1421  
 RESULT 10  
 CAL4\_MOUSE  
 ID CAL4\_MOUSE STANDARD; PRT; 1669 AA.  
 AC P02463;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Collagen alpha 1(IV) chain precursor.  
 GN COL4A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89197932; PubMed=2703490;  
 RA Muthukumar G., Blumberg B., Kurkinen M.;  
 RT "The complete primary structure for the alpha 1-chain of mouse  
 RT collagen IV. Differential evolution of collagen IV domains.";  
 RL J. Biol. Chem. 264:6310-6317(1989).  
 RN [2]  
 RP SEQUENCE OF 1-1154 FROM N.A.  
 RX MEDLINE=88112221; PubMed=3338568;  
 RA Wood L., Theriault N., Vogeli G.;  
 RT "CDNA clones completing the nucleotide and derived amino acid  
 RT sequence of the alpha 1 chain of basement membrane (type IV) collagen  
 RT from mouse.";  
 RL FEBS Lett. 227:5-8(1988).  
 RN [3]  
 RP SEQUENCE OF 1149-1424 FROM N.A.  
 RX MEDLINE=86301886; PubMed=3755692;  
 RA Nath P., Laurent M., Horn E., Sobel M.E., Zon G., Vogeli G.;  
 RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a  
 RT synthetic oligodeoxynucleotide.";  
 RL Gene 43:301-304(1986).  
 RN [4]  
 RP SEQUENCE OF 1276-1669 FROM N.A.

RX MEDLINE=85127033; PubMed=2578961;  
RA Oberbauer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,  
RA Vogeli G., Voss T., Siebold B., Glanville R.W., Kuhn K.;  
RT "Amino acid sequence of the non-collagenous globular domain (NC1) of  
RT the alpha 1(IV) chain of basement membrane collagen as derived from  
RT complementary DNA.";  
RL Eur. J. Biochem. 147:217-224 (1985).  
RN [5]  
RP SEQUENCE OF 1441-1669 FROM N.A.  
RX MEDLINE=87250460; PubMed=3597389;  
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,  
RA Saus J., Pihlajaniemi T.;  
RT "Extensive homology between the carboxyl-terminal peptides of mouse  
RT alpha 1(IV) and alpha 2(IV) collagen.";  
RL J. Biol. Chem. 262:8496-8499 (1987).  
RN [6]  
RP PARTIAL SEQUENCE FROM N.A.  
RX MEDLINE=86196099; PubMed=3009469;  
RA Sakurai Y., Sullivan M., Yamada Y.;  
RT "Alpha 1 type IV collagen gene evolved differently from fibrillar  
RT collagen genes.";  
RL J. Biol. Chem. 261:6654-6657 (1986).  
RN [7]  
RP SEQUENCE OF 1-28 FROM N.A.  
RX MEDLINE=89066738; PubMed=3198626;  
RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;  
RT "Head-to-head arrangement of murine type IV collagen genes.";  
RL J. Biol. Chem. 263:19274-19277 (1988).  
RN [8]  
RP SEQUENCE OF 1-28 FROM N.A.  
RX MEDLINE=89071759; PubMed=3200851;  
RA Burdello P.D., Martin G.R., Yamada Y.;  
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a  
RT bidirectional promoter and a shared enhancer.";  
RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682 (1988).  
RN [9]  
RP SEQUENCE OF 1-129 FROM N.A.  
RX MEDLINE=88243724; PubMed=3379041;  
RA Killen P.D., Burdello P., Sakurai Y., Yamada Y.;  
RT "Structure of the amino-terminal portion of the murine alpha 1(IV)  
RT collagen chain and the corresponding region of the gene.";  
RL J. Biol. Chem. 263:8706-8709 (1988).  
CC -!- FUNCTION: Type IV collagen is the major structural component of  
CC glomerular basement membranes (GBM), forming a 'chicken-wire'  
CC meshwork together with laminins, proteoglycans and entactin/  
CC nidogen.  
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -  
CC alpha 6(IV), each of which can form a triple helix structure with  
CC 2 other chains to generate type IV collagen network.  
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
CC domain (NC1) at their C-terminus, frequent interruptions of the G-  
CC X-Y repeats in the long central triple-helical domain (which may  
CC cause flexibility in the triple helix), and a short N-terminal  
CC triple-helical 7S domain.  
CC -!- PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
CC are involved in inter- and intramolecular disulfide bonding. 12 of  
CC these, located in the NC1 domain, are conserved in all known type  
CC IV collagens.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; J03758; AAA37439.1; -  
DR EMBL; M23333; AAA51625.1; -  
DR EMBL; J04694; AAA50292.1; -  
DR EMBL; X06777; CAA29946.1; -

DR EMBL; X02201; CAA26132.1; -  
DR EMBL; M15832; AAA37340.1; -  
DR EMBL; M14042; AAA37342.1; -  
DR EMBL; M12879; AAA37343.1; -  
DR EMBL; M13024; -; NOT ANNOTATED CDS.  
DR EMBL; M13025; -; NOT ANNOTATED CDS.  
DR EMBL; M13026; AAA37344.1; -  
DR EMBL; M13027; AAA37345.1; -  
DR EMBL; M13043; AAA37346.1; -  
DR EMBL; J04448; AAA37437.1; -  
DR PIR; A33525; CGMS4B.  
DR MGD; MGI:88454; Col4a1.  
DR GO; GO:0005604; C:basement membrane; IDA.  
DR InterPro; IPR008161; C1q helix.  
DR InterPro; IPR008160; Collagen.  
DR Pfam; PF01413; C4; 2.  
DR ProDom; PD000007; C1q helix; 6.  
DR SMART; SM00111; C4; 2.  
KW Extracellular matrix; Connective tissue; Basement membrane;  
KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.  
FT SIGNAL 1 27  
FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).  
FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.  
FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.  
FT DOMAIN 1441 1669 NONHELICAL REGION (NC1).  
FT DISULFID 1460 1551 OR 1548 (BY SIMILARITY).  
FT DISULFID 1493 1548 OR 1551 (BY SIMILARITY).  
FT DISULFID 1505 1511 BY SIMILARITY.  
FT DISULFID 1570 1665 OR 1662 (BY SIMILARITY).  
FT DISULFID 1604 1662 OR 1665 (BY SIMILARITY).  
FT DISULFID 1616 1622 BY SIMILARITY.  
FT CARBOHYD 126 126 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CONFLICT 26 26 A -> P (IN REF. 2).  
FT CONFLICT 186 186 S -> L (IN REF. 2).  
FT CONFLICT 319 319 Q -> S (IN REF. 2).  
FT CONFLICT 369 369 Q -> L (IN REF. 2).  
FT CONFLICT 403 403 L -> F (IN REF. 2).  
FT CONFLICT 481 481 P -> L (IN REF. 2).  
FT CONFLICT 493 493 Q -> H (IN REF. 2).  
FT CONFLICT 712 712 Q -> I (IN REF. 2).  
FT CONFLICT 813 813 E -> Q (IN REF. 2).  
FT CONFLICT 982 982 Q -> H (IN REF. 2).  
FT CONFLICT 1397 1397 V -> S (IN REF. 3).  
SQ SEQUENCE 1669 AA; 160680 MW; 42916B91E52058E9 CRC64;  
  
Query Match 46.4%; Score 501; DB 1; Length 1669;  
Best Local Similarity 46.2%; Pred. No. 3.9e-22;  
Matches 108; Conservative 14; Mismatches 64; Indels 48; Gaps 7;  
  
QY 1 GIRLKGTKGKGBDGFPGFKDNGI---KGDREIGPPGRGBDG-----PEGPKG 49  
Db 1177 GPPSKGDKSGKEVGFPGFAGSPGIPGVKGEQGMGPPGQPGPLGTGHPVGEKPG 1236  
QY 50 RGGNGD-----PGPLGP-----GKGLGVPLGVPGRGQPKSGIGFPGF-- 92  
Db 1237 DRGQGGQGLGHPGMPGPPGPGINGPKGDKNGQWPGAVGVPKPGDQGMPIGG 1296  
QY 93 -PGANGKGRGTGKPGPRGQRTGPRGERGRGITGKPGKNSGGDPAGP----- 146  
Db 1297 SFGITGSKDMLGPGVPGFQKGLGVLGVKQDQGVGPGKLGQPPGPGYDVIK 1356  
QY 147 --PBERGNGQGRTGPPGPKGP-----PGPKDGLPGHPGQRTG 186  
Db 1357 GEPLGPGEPGLKGLGPPGPGQGVTSVGLPGPPGPGVPGFDGAPGQGET 1410  
  
RESULT 11  
CALA\_CHICK  
ID CALA\_CHICK  
AC P08125; STANDARD; PRT; 674 AA.















**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 13:05:48 ; Search time 20.8274 Seconds  
(without alignments)  
2817.753 Million cell updates/sec

Title: US-09-775-964-6  
Perfect score: 1079  
Sequence: 1 GRLGLKGTGKGEDGPGF.....PPGPPKGLGPHGQGRGET 186

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1079	100.0	1838	4 Q15094	Q15094 homo sapien
2	1071	99.3	1838	11 O88207	O88207 mus musculus
3	1071	99.3	1840	11 Q60467	Q60467 cricetus
4	1071	99.3	1840	11 Q9J103	Q9J103 rattus norv
5	1037	96.1	1835	13 Q91AU4	Q91AU4 gallus gall
6	945	87.6	1804	11 Q8QWR4	Q8QWR4 mus musculus
7	885	82.0	1820	13 Q91907	Q91907 pagrus majo
8	880	81.6	1827	13 Q8UUM5	Q8UUM5 oryzias lat
9	741	68.7	1053	11 Q920S4	Q920S4 mus musculus
10	741	68.7	1739	11 Q9JL12	Q9JL12 mus musculus
11	723	67.0	1737	11 Q9J104	Q9J104 rattus norv
12	533.5	49.4	1349	13 Q8AW17	Q8AW17 brachydanio
13	509.5	47.2	890	5 Q77087	Q77087 alvinella p
14	506.5	46.9	1860	4 Q81SC6	Q81SC6 homo sapien
15	501.5	46.5	886	4 Q9NUB7	Q9NUB7 homo sapien
16	497.5	46.1	888	13 Q90796	Q90796 gallus gall

ALIGNMENTS

RESULT 1

Q15094 PRELIMINARY; PRT; 1838 AA.

AC Q15094;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Pro-alpha-1 type V collagen.  
GN COL5A1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92105142; PubMed=172213;  
RA Greenspan D.S., Cheng W., Hoffman G.G.;  
RT "The pro-alpha 1(V) collagen chain. Complete primary structure,  
RT distribution of expression, and comparison with the pro-alpha 1(XI)  
RT collagen chain.";  
RL J. Biol. Chem. 266:24727-24733 (1991).  
DR EMBL; M76729; AAA5993.1; -;  
DR PIR; S18802; CGHUIV.  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR GO; GO:0007155; P:cell adhesion; IEA.  
DR InterPro; IPR008161; Clg helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR008985; ConA like lec gl.  
DR InterPro; IPR000885; Fib.Collagen\_C.  
DR InterPro; IPR001791; Laminin\_G.  
DR InterPro; IPR003129; TSPN.  
DR Pfam; PF01410; COLFI; 1.  
DR Pfam; PF01391; Collagen; 17.  
DR Pfam; PF02210; TSPN; 1.  
DR ProDom; PD000007; Clg helix; 1.  
DR ProDom; PD002078; Fib.collagen\_C; 1.  
DR SMART; SM00038; COLFI; 1.  
DR SMART; SM00282; LamG; 1.

```
DR SMART; SM00210; TSPN; 1.
KW COLLAGEN.
SQ SEQUENCE 1838 AA; 183610 MW; 5078307F6E00F0BA CRC64;

Query Match 100.0%; Score 1079; DB 4; Length 1838;
Best Local Similarity 100.0%; Pred. No. 8e-79;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GIRLKGTKGKEDGFPFGKDMGIKGRGEIGPRGCEGDGEGPKRGCGNGDPGPL 60
DB 799 GIRLKGTKGKEDGFPFGKDMGIKGRGEIGPRGCEGDGEGPKRGCGNGDPGPL 858

QY 61 GPPGEKGLGVPLPGYPRGQKSGIGFPPGANGKGGRTGPKGPRGQGTGPR 120
DB 859 GPPGEKGLGVPLPGYPRGQKSGIGFPPGANGKGGRTGPKGPRGQGTGPR 918

QY 121 GERGPRGITKPGPKNSGGDGAGPPGERGPNPGQGTGFPKGPFGPKDGLPGHP 180
DB 919 GERGPRGITKPGPKNSGGDGAGPPGERGPNPGQGTGFPKGPFGPKDGLPGHP 978

QY 181 QORGET 186
DB 979 QORGET 984

RESULT 2
O88207 PRELIMINARY; PRT; 1838 AA.
AC O88207;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE COLLAGEN a1(V).
GN COL5A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98250615; PubMed=9582436;
RA Wu Y.L., Sumiyoshi H., Khaleuzzaman M., Ninomiya Y., Yoshioka H.;
RT "CDNA sequence and expression of the mouse alpha1(V) collagen gene.";
RL Biochim. Biophys. Acta 1397:275-284(1998).
DR EMBL; AB009993; BAA28786.1; -.
DR MGD; MGI:88457; Col5a1.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Clg_helix; 2.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
KW Collagen.
SQ SEQUENCE 1838 AA; 183691 MW; 7A520B23D1851783 CRC64;

Query Match 99.3%; Score 1071; DB 11; Length 1838;
Best Local Similarity 99.5%; Pred. No. 3.5e-78;
Matches 185; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GIRLKGTKGKEDGFPFGKDMGIKGRGEIGPRGCEGDGEGPKRGCGNGDPGPL 60
DB 799 GIRLKGTKGKEDGFPFGKDMGIKGRGEIGPRGCEGDGEGPKRGCGNGDPGPL 858
```

```
QY 61 GPPGEKGLGVPLPGYPRGQKSGIGFPPGANGKGGRTGPKGPRGQGTGPR 120
DB 859 GPPGEKGLGVPLPGYPRGQKSGIGFPPGANGKGGRTGPKGPRGQGTGPR 918

QY 121 GERGPRGITKPGPKNSGGDGAGPPGERGPNPGQGTGFPKGPFGPKDGLPGHP 180
DB 919 GERGPRGITKPGPKNSGGDGAGPPGERGPNPGQGTGFPKGPFGPKDGLPGHP 978

QY 181 QORGET 186
DB 979 QORGET 984

RESULT 3
Q60467 PRELIMINARY; PRT; 1840 AA.
AC Q60467;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Pro-alpha-1 type V collagen.
OS Cricetulus longicaudatus (Long-tailed hamster) (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10030;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92105142; PubMed=1722213;
RA Greenspan D.S., Cheng W., Hoffman G.G.;
RT "The pro-alpha-1(V) collagen chain: Complete primary structure,
RT distribution of expression, and comparison with the pro-alpha-1(XI)
RT collagen chain.";
RL J. Biol. Chem. 266:24727-24733(1991).
DR EMBL; M76730; AAA37002.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Clg_helix; 2.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
KW Collagen.
SQ SEQUENCE 1840 AA; 184174 MW; 32C56821EF64CE75 CRC64;

Query Match 99.3%; Score 1071; DB 11; Length 1840;
Best Local Similarity 99.5%; Pred. No. 3.6e-78;
Matches 185; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GIRLKGTKGKEDGFPFGKDMGIKGRGEIGPRGCEGDGEGPKRGCGNGDPGPL 60
DB 801 GIRLKGTKGKEDGFPFGKDMGIKGRGEIGPRGCEGDGEGPKRGCGNGDPGPL 860

QY 61 GPPGEKGLGVPLPGYPRGQKSGIGFPPGANGKGGRTGPKGPRGQGTGPR 120
DB 861 GPPGEKGLGVPLPGYPRGQKSGIGFPPGANGKGGRTGPKGPRGQGTGPR 920

QY 121 GERGPRGITKPGPKNSGGDGAGPPGERGPNPGQGTGFPKGPFGPKDGLPGHP 180
DB 921 GERGPRGITKPGPKNSGGDGAGPPGERGPNPGQGTGFPKGPFGPKDGLPGHP 980

QY 181 QORGET 186
```

```
Db          981 GORGET 986
|||||
RESULT 4
Q9J103      PRELIMINARY;      PRT; 1840 AA.
AC Q9J103
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha 1 type V collagen.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;
[1]
SEQUENCE FROM N.A.
RX STRAIN=Sprague Dawley;
RX MEDLINE=20428740; PubMed=10852920;
RA Chernousov M.A., Rothblum K., Tyler W.A., Stahl R.C., Carey D.J.;
RT "Schwann cells synthesize type V collagen that contains a novel alpha
RT 4 chain. molecular cloning, biochemical characterization, and high
RT affinity heparin binding of alpha4(V) collagen.";
RL J. Biol. Chem. 275:28208-28215(2000).
DR ENBL; AF272662; AAF76433.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008985; ConA_like_lect_g1.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR00508; Peptidase_S26.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Clg_helix; 2.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
DR PROSITE; PS00761; SPASE_I_3; 1.
KW Collagen.
SQ SEQUENCE 1840 AA; 183986 MW; AD38F5FF8686923C CRC64;

Query Match          99.3%; Score 1071; DB 11; Length 1840;
Best Local Similarity 99.5%; Pred. No. 3,6e-78;
Matches 185; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GIRLKGTKGKEDGPGFGKDMGKGRGEIGPPGPRGEDGPGKRGSGNGDPGPL 60
Db 801 GIRLKGTKGKEDGPGFGKDMGKGRGEIGPPGPRGEDGPGKRGSGNGDPGPL 860

QY 61 GPPGKGLGVPLPGVPGQPKSGISGFFPGANGKGGRTGKPGPRGQGTGPR 120
Db 861 GPTGKGLGVPLPGVPGQPKSGISGFFPGANGKGGRTGKPGPRGQGTGPR 920

QY 121 GERGPGRITGKPGKNGSGDGPAGPPGPRGNGPQGTGTPGPKPPGPKDGLPGHP 180
Db 921 GERGPGRITGKPGKNGSGDGPAGPPGPRGNGPQGTGTPGPKPPGPKDGLPGHP 980

QY 181 GORGET 186
|||||
Db 981 GORGET 986

RESULT 5
```

```
Q9IAU4      PRELIMINARY;      PRT; 1835 AA.
AC Q9IAU4
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha 1 (V) collagen.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
NCBI_TaxID=9031;
[1]
SEQUENCE FROM N.A.
RX STRAIN=bred White Leghorn;
RX MEDLINE=20068042; PubMed=10601735;
RA Gordon M.K., Marchant J.K., Foley J.W., Igoe F., Gibney E.P.,
RA Nah H.D., Barembaum M., Myers J.C., Rodriguez E., Dublet B.,
RA van der Rest M., Linsenmayer T.F., Upholt W.B., Birk D.E.;
RT "Complete primary structure of the chicken alpha1(V) collagen chain.";
RL Matrix Biol. 18:481-486(1999).
DR PIR; A55047; A55047.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008985; ConA_like_lect_g1.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 19.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Clg_helix; 2.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
KW Collagen.
SQ SEQUENCE 1835 AA; 184234 MW; D05B9D71022D4B2 CRC64;

Query Match          96.1%; Score 1037; DB 13; Length 1835;
Best Local Similarity 96.2%; Pred. No. 2e-75;
Matches 179; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 GIRLKGTKGKEDGPGFGKDMGKGRGEIGPPGPRGEDGPGKRGSGNGDPGPL 60
Db 795 GVRGLKGTKGKEDGPGFGKDMGKGRGEIGPPGPRGEDGPGKRGSGNGDPGPL 854

QY 61 GPPGKGLGVPLPGVPGQPKSGISGFFPGANGKGGRTGKPGPRGQGTGPR 120
Db 855 GPAGEKGLGVPLPGVPGQPKSGISGFFPGANGKGGRTGKPGPRGQGTGPR 914

QY 121 GERGPGRITGKPGKNGSGDGPAGPPGPRGNGPQGTGTPGPKPPGPKDGLPGHP 180
Db 915 GERGPGRITGKPGKNGSGDGPAGPPGPRGNGPQGTGTPGPKPPGPKDGLPGHP 974

QY 181 GORGET 186
|||||
Db 975 GORGET 980

RESULT 6
Q80WR4      PRELIMINARY;      PRT; 1804 AA.
AC Q80WR4;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collial protein.
OS Mus musculus (Mouse).
```





OS Oryzias latipes (Medaka fish) (Japanese ricefish).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
OC Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.  
OX NCBI\_TaxID=8090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Hd-rR;  
RA Matsuo M.Y., Asakawa S., Shimizu N., Kimura H., Nonaka M.;  
RT "Nucleotide Sequence of the MHC Class I Region of a Teleost, the  
Medaka";  
RL Immunogenetics 0:0-0(2002).  
DR EMBL; AB073376; BAB83839.1; -.  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR GO; GO:0007155; P:cell adhesion; IEA.  
DR InterPro; IPR008161; Clg\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR008985; ConA\_like\_lec\_gl.  
DR InterPro; IPR001791; Laminin\_C.  
DR InterPro; IPR003129; TSPN.  
DR Pfam; PF01410; COLFI; 1.  
DR Pfam; PF01391; Collagen; 17.  
DR Pfam; PF02210; TSPN; 1.  
DR ProDom; PD000007; Clg\_helix; 3.  
DR ProDom; PD002078; Fib\_collagen\_C; 1.  
DR SMART; SM00038; COLFI; 1.  
DR SMART; SM00282; LamG; 1.  
DR SMART; SM00210; TSPN; 1.  
KW Collagen.  
SQ SEQUENCE 1827 AA; 181589 MW; AF5C71E9B38906EA CRC64;  
  
Query Match 81.6%; Score 880; DB 13; Length 1827;  
Best Local Similarity 83.2%; Pred. No. 8.9e-63;  
Matches 154; Conservative 7; Mismatches 24; Indels 0; Gaps 0;  
  
QY 1 GIRLKGTKGKEDGPGFKGDMGKIGDRGEIGPPGPRGEDGPGKGGGNGDPGL 60  
DB 807 GIRLKGKHKGEKGDPGFKGDMGKIGDRGEIGPPGPRGEDGPGKGGGNGDPGL 866  
  
QY 61 GPPGEGKGLVPGLPVPGVPGKQSGISGFFPGANGKGGKRGTPGKPRGQGTGPR 120  
DB 867 GTVGEKGLVPGLPVPGVPGKQSGISGFFPGANGKGGKRGTPGKPRGQGTGPR 926  
  
QY 121 GERGPRGITKPGKNGSGDGPAGPPGERGNGPQGTGFPKGPDPGKGLPGHP 180  
DB 927 GQGRPGATGKAGKGTSGSDGPPPPGERGLPQANGFPKGPDPGKGLPGHP 986  
  
QY 181 GQGRGE 185  
DB 987 GQGRGE 991  
  
RESULT 9  
Q920S4 PRELIMINARY; PRT; 1053 AA.  
AC Q920S4;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Adipocyte-specific protein 6.  
GN COL5A3 OR ASP6.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Tsuruga H.;  
RT "Adipocyte-specific protein 6, a novel protein upregulated during  
adipocyte differentiation.";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB040491; BAB68504.1; -.  
DR MGD; MGI:1858212; Col5a3.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0007155; P:cell adhesion; IEA.  
DR InterPro; IPR008161; Clg\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR008985; ConA\_like\_lec\_gl.  
DR InterPro; IPR001791; Laminin\_G.  
DR Pfam; PF01391; Collagen; 10.  
DR Pfam; PF02210; TSPN; 1.  
DR ProDom; PD000007; Clg\_helix; 3.  
DR SMART; SM00282; LamG; 1.  
DR SMART; SM00210; TSPN; 1.  
DR PROSITE; PS50025; LAM\_G\_DOMAIN; 1.  
KW Collagen.  
SQ SEQUENCE 1053 AA; 104534 MW; C220E48BB8D1F6CC CRC64;  
  
Query Match 68.7%; Score 741; DB 11; Length 1053;  
Best Local Similarity 73.0%; Pred. No. 8.4e-52;  
Matches 135; Conservative 10; Mismatches 40; Indels 0; Gaps 0;  
  
QY 1 GIRLKGTKGKEDGPGFKGDMGKIGDRGEIGPPGPRGEDGPGKGGGNGDPGL 60  
DB 718 GNRGLQGEKGEKGEDGPGFKGDMGKIGDRGEIGPPGPRGEDGPGKGGGNGDPGL 777  
  
QY 61 GPPGEGKGLVPGLPVPGVPGKQSGISGFFPGANGKGGKRGTPGKPRGQGTGPR 120  
DB 778 GAAGEKGLVPGLPVPGVPGKQSGISGFFPGANGKGGKRGTPGKPRGQGTGPR 837  
  
QY 121 GERGPRGITKPGKNGSGDGPAGPPGERGNGPQGTGFPKGPDPGKGLPGHP 180  
DB 838 GDRGPGATGQPGKGDVGVNGSGPPGKGLPGLQGPFGPKGPPGQKDGISGHP 897  
  
QY 181 GQGRGE 185  
DB 898 GQGRGE 902  
  
RESULT 10  
Q9JLI2 PRELIMINARY; PRT; 1739 AA.  
AC Q9JLI2;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Collagen type V alpha 3 chain.  
GN COL5A3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=20187594; PubMed=10722718;  
RA Imamura Y., Scott I.C., Greenspan D.S.;  
RT "The pro-alpha3 (V) collagen chain: Complete primary structure,  
expression domains in adult and developing tissues, and comparison to  
the structures and expression domains of the other types V and XI  
procollagen chains.";  
RL J. Biol. Chem. 275:8749-8759(2000).  
DR EMBL; AF176645; AAF59901.1; -.  
DR MGD; MGI:1858212; Col5a3.  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR GO; GO:0007155; P:cell adhesion; IEA.  
DR InterPro; IPR008161; Clg\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR008985; ConA\_like\_lec\_gl.  
DR InterPro; IPR000885; Fib\_collagen\_C.  
DR InterPro; IPR001791; Laminin\_G.

DR InterPro; IPR003129; TSPN.  
DR Pfam; PF01410; COLFI; 1.  
DR Pfam; PF01391; Collagen; 16.  
DR Pfam; PF02210; TSPN; 1.  
DR ProDom; PD000007; Clg helix; 5.  
DR ProDom; PD002078; Fib\_collagen\_C; 1.  
DR SMART; SM00038; COLFI; 1.  
DR SMART; SM00282; LamG; 1.  
DR SMART; SM00210; TSPN; 1.  
DR PROSITE; PS50025; LAM\_G\_DOMAIN; 1.  
KW Collagen.  
SQ SEQUENCE 1739 AA; 171968 MW; EFDDEBA46E59984 CRC64;  
  
Query Match 68.7%; Score 741; DB 11; Length 1739;  
Best Local Similarity 73.0%; Pred. No. 1.4e-51;  
Matches 135; Conservative 10; Mismatches 40; Indels 0; Gaps 0;  
  
QY 1 GIRGLKGTGKEGDEGPFPGKGMGKIDRGEIGPPGPRGDEGPEGPKRGPGNGDGPGL 60  
DB 718 GNRGLQGEKGERGEDGPFPGKDGKDRGNPGLPGRGDEGPEGQKGFGLPDEGPP 777  
  
QY 61 GPPGEGKGLGVPLGPGYPRGQPGKSGIGPPGPGANGKGGRTGCKPGPRGQGTGPR 120  
DB 778 GAAGEKGLGVPLGPGYPRGQPGKSGIGPPGPGANGKGGRTGCKPGPRGQGTGPR 837  
  
QY 121 GERGPRGITGKPGPKNSGGDGPAGPPGERGPNPGQGTGFPKPGPPGPKDGLPGHP 180  
DB 838 GDRGQPGATGQPGKGDVGQNGSPGPFGEKGLPLGLQPGFPGPKPGPPQKDGISGHP 897  
  
QY 181 GQGE 185  
DB 898 GQGE 902

## RESULT 11

Q9J104 PRELIMINARY; PRT; 1737 AA.  
AC Q9J104;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Alpha 4 type V collagen.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Sprague Dawley;  
RX MEDLINE=20428740; PubMed=10852920;  
RA Chernousov M.A., Rothblum K., Tyler W.A., Stahl R.C., Carey D.J.;  
RT "Schwann cells synthesize type V collagen that contains a novel alpha  
RT 4 chain. molecular cloning, biochemical characterization, and high  
RT affinity heparin binding of alpha4(V) collagen.";  
RL J. Biol. Chem. 275:28208-28215(2000).  
DR EMBL; AF272661; AAF76432.1; -  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR GO; GO:0007155; P:cell adhesion; IEA.  
DR InterPro; IPR008161; Clg helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR008985; ConA like lec.gl.  
DR InterPro; IPR008885; Fib\_collagen\_C.  
DR InterPro; IPR001791; Laminin\_G.  
DR InterPro; IPR003129; TSPN.  
DR Pfam; PF01410; COLFI; 1.  
DR Pfam; PF01391; Collagen; 17.  
DR Pfam; PF02210; TSPN; 1.  
DR ProDom; PD000007; Clg helix; 3.  
DR ProDom; PD002078; Fib\_collagen\_C; 1.  
DR SMART; SM00038; COLFI; 1.  
DR SMART; SM00282; LamG; 1.  
DR SMART; SM00210; TSPN; 1.

KW Collagen.  
SQ SEQUENCE 1737 AA; 171574 MW; D635D5D57481C257 CRC64;  
  
Query Match 67.0%; Score 723; DB 11; Length 1737;  
Best Local Similarity 71.9%; Pred. No. 3.9e-50;  
Matches 133; Conservative 10; Mismatches 42; Indels 0; Gaps 0;  
  
QY 1 GIRGLKGTGKEGDEGPFPGKGMGKIDRGEIGPPGPRGDEGPEGPKRGPGNGDGPGL 60  
DB 716 GNRGLQGEKGERGEDGPFPGKDGKDRGNPGLPGRGDEGPEGQKGFGLPDEGPP 775  
  
QY 61 GPPGEGKGLGVPLGPGYPRGQPGKSGIGPPGPGANGKGGRTGCKPGPRGQGTGPR 120  
DB 776 GAAGEKGLGVPLGPGYPRGQPGKSGIGPPGPGANGKGGRTGCKPGPRGQGTGPR 835  
  
QY 121 GERGPRGITGKPGPKNSGGDGPAGPPGERGPNPGQGTGFPKPGPPGPKDGLPGHP 180  
DB 836 GDRGQPGATGQPGKGDVGQNGSPGAPGEKGLPLQGGPPGPKPGPPQKDGIPGHP 895  
  
QY 181 GQGE 185  
DB 896 GQGE 900

## RESULT 12

Q9AW17 PRELIMINARY; PRT; 1349 AA.  
AC Q9AW17;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE SI:GZ12F11.3 (Collagen type XI alpha-2 (Fragment)).  
GN COL11A2.  
OS Brachydanio rerio (Zebrafish) (Danio rerio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Skuce C.;  
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL672176; CAD54661.1; -  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR GO; GO:0007155; P:cell adhesion; IEA.  
DR InterPro; IPR008161; Clg helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR008985; ConA like lec.gl.  
DR InterPro; IPR008885; Fib\_collagen\_C.  
DR InterPro; IPR001791; Laminin\_G.  
DR InterPro; IPR003129; TSPN.  
DR Pfam; PF01410; COLFI; 1.  
DR Pfam; PF01391; Collagen; 12.  
DR Pfam; PF02210; TSPN; 1.  
DR ProDom; PD000007; Clg helix; 2.  
DR ProDom; PD002078; Fib\_collagen\_C; 1.  
DR SMART; SM00038; COLFI; 1.  
DR SMART; SM00282; LamG; 1.  
DR SMART; SM00210; TSPN; 1.  
KW Collagen.  
FT NON TER  
SQ SEQUENCE 1349 AA; 135634 MW; 1DD0432D7B63BF9C CRC64;

Query Match 49.4%; Score 533.5; DB 13; Length 1349;  
Best Local Similarity 43.1%; Pred. No. 5.8e-35;  
Matches 115; Conservative 14; Mismatches 55; Indels 83; Gaps 7;  
  
QY 1 GIRGLKGTGKEGDEGPFPGKGMGKIDRGEI----- 33  
DB 389 GIRGLKHGKGEDEGPFPGKGDGVKGERKGLGVPGIPGYPGRGKIGLTGPQGANFPP 448  
  
QY 34 ---GPPGPRGED-----GPEPKRGGPN-----GDPGLG 61

```
Db 449 GPKGPPGPKDGLGPHGPGQGEFGQKGVPPGPPGVVGHPSGSGTGQMGGRHHPGP 508
Qy 62 PPGE-----KGKLGVPFLPGYPGRGKSGISF-----PGFPGANGKGGRTGP 106
Db 509 PPGEQGLSGSGKGTGKDGPPGPGKDGPPGLRGPPGGRGLPGTGPSSGLKGNBPAG 568
Qy 107 KPGRGPGRTGPRGGRGRTGKPKPKNSGSDGAGPPGGRGNP-----QGPTG 160
Db 569 PPGPAGSSGGRGAGTAGTAPVGPFGPPGPGTSGEKVGEKGPVGPAGRDGIQGVG 628
Qy 161 FPGPKGPPGPKD---GLPCHPGQRG 184
Db 629 LPGPAGPPGSGEDGKGEVGEPCQG 655

RESULT 13
ID O77087 PRELIMINARY; PRT; 890 AA.
AC O77087;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Fibrillar collagen chain Fap1 alpha.
OS Alvinella pompejana.
OC Eukaryota; Metazoa; Annelida; Polychaeta; Palpata; Canalipalpata;
OC Terebellida; Alvinellidae; Alvinella.
ON NCBI_TaxID=6376;
RN [1]
RP SEQUENCE FROM N.A.
RA Sicot F.-X., Gaill F., Exposito J.-Y., Garrone R., Deutsch J.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR GO: AF053538; AAC35289.2; -.
DR GO: GO:0005581; C:collagen; IEA.
DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro: IPR008161; Clq helix.
DR InterPro: IPR008160; Collagen.
DR InterPro: IPR00885; Fib_collagen_C.
DR Pfam: PF01410; COLFI; 1.
DR ProDom: PD00007; Clq helix; 9.
DR ProDom: PD002078; Fib_collagen_C; 1.
DR SMART: SM00038; COLFI; 1.
KW Collagen.
SQ SEQUENCE 890 AA; 88137 MW; 52CC756FA70CA90C CRC64;

Query Match 47.2%; Score 509.5; DB 5; Length 890;
Best Local Similarity 47.5%; Pred. No. 3.3e-33;
Matches 106; Conservative 14; Mismatches 64; Indels 39; Gaps 5;

Qy 1 GIRLKGTKGKEDGPPGPKGDM---GIKGDREIGPPGPRGEDGP---EGPKGRGGPN 54
Db 355 GTFGMDTKGERGEDGPPGLPGTGPFGPQGERGVGLFGERGEPFGAPGRGFGAP 414
Qy 55 GDRGPIGPGEKGLGVPLGVPYVGRQKSGTGFPGFPGANGKGGRTGPKGP-----110
Db 415 GEPGDTGPRGSGSPGAPGAPGAAGPCCGPKDGLPLAGRPDKGPPGPPGGLA 474
Qy 111 -----RGQGTGPRGERGRGITGKPGKNSGGGDPAGPP-----GERGNGQP 156
Db 475 GAPFMQLGPPGPTGFGGERGERGFGAPGVGPPGPPGAPAGQQGSKGRGAAGPK 534
Qy 157 GPTGFP-----GPKGPPGPKDGLPCHPGQRG 184
Db 535 GDKGWFPMPLQGMQGTGGERGEPGPPGPPGADGMKRGEGSG 577

RESULT 14
ID O81ZC6 PRELIMINARY; PRT; 1860 AA.
AC O81ZC6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
```

```
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen XXVII proalpha 1 chain precursor.
GN COL27A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Pace J.M., Corrado M., Missero C., Byers P.H.;
RT Tissue=Cartilage;
RT "Identification, characterization and expression analysis of a new
RT fibrillar collagen gene, COL27A1.";
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY149237; AA41263.1; -.
DR GO: GO:0005581; C:collagen; IEA.
DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO: GO:0007155; P:cell adhesion; IEA.
DR InterPro: IPR008160; Collagen.
DR InterPro: IPR00885; ConA like lec_gl.
DR InterPro: IPR00885; Fib_collagen_C.
DR InterPro: IPR003129; TSPN.
DR Pfam: PF01410; COLFI; 1.
DR Pfam: PF01391; Collagen; 15.
DR Pfam: PF02210; TSPN; 1.
DR ProDom: PD002078; Fib_collagen_C; 1.
DR SMART: SM00038; COLFI; 1.
DR SMART: SM00210; TSPN; 1.
KW Signal.
FT SIGNAL.
FT CHAIN 1 41 POTENTIAL.
FT CHAIN 625 1621 COLLAGEN XXVII PROALPHA 1 CHAIN.
SQ SEQUENCE 1860 AA; 186891 MW; 5F8CDFAF4B6014EC CRC64;

Query Match 46.9%; Score 506.5; DB 4; Length 1860;
Best Local Similarity 36.4%; Pred. No. 1.2e-32;
Matches 121; Conservative 20; Mismatches 44; Indels 147; Gaps 13;

Qy 1 GIRLKGTK-----KGEKEDGPPGPKGDMGKNGRGE-----32
Db 838 GMKGLMGSVGEPGLKGDGEGVGVSGDPFGQDKGSQGLPGFPGARGKPGPLGKVGDK 897
Qy 33 -----IGPPGPR-----GSDGPE-----GPKRGSGPNDGPPGPGP-----62
Db 898 GSIGFEGPFGGPGFDIGPNDGNGPEGMKRGKPGARGLFPGRQLGPEGDEGPMGPPGAP 957
Qy 63 -----PGEKGLGVPL-----PGYVPGRGKSGTGFPGF-----PGANGK-----99
Db 958 GLEGQDGRKGFPGPGLDGVKGEPPGDPGRPPVGEQGMGFGTGLVGEPIVGEKGRGMM 1017
Qy 100 -----GGRGTGPKGP-----RGQGTGPKGP-----123
Db 1018 GPPGVPGPKSGMHPGMPGCMGTGPGPGQPGSGRPGMRGAKGRRGPRGPDGPAGEQ 1077
Qy 124 GPRGITGPKPKNSG-----GD-----GPAGP-----PGRGNGQP 156
Db 1078 GSRGLKGPGRPCRPGRPCQQVAGRGHLGSRGFPFGIPGSPGPGTGLGPGEPGQGP 1137
Qy 157 GPTGEP---GPKGPPGPKDGLPCHPGQGE 185
Db 1138 GPFGPPGEMGPKGPPGAVGEPGLFGEAGMKGD 1169

RESULT 15
ID Q9NUB7 PRELIMINARY; PRT; 886 AA.
AC Q9NUB7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE DA24A23.1 (Collagen, type IV, alpha 5 (Alport syndrome))
DE (Fragment).
GN COL4A5.
OS Homo sapiens (Human).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Codley V.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035425; CAB90289.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD000007; Clg_helix; 1.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON TER
SQ SEQUENCE      1      1      85479 MW; 8C06B9FCA9AA6569 CRC64;

Query Match      46.5%; Score 501.5; DB 4; Length 886;
Best Local Similarity 47.2%; Pred. No. 1.5e-32;
Matches 108; Conservative 13; Mismatches 63; Indels 45; Gaps 7;

QY      1 GIRGLKGTGKGEKGEDGPPGPKGDMGIKGDGE-----IGPPGPRGEDGP--EGPKG 49
Db      394 GPFGLPGLSQKGDGGLPGIPGNGLPGPKGEFGFGVQVQGPVPGSPGPALEGPKG 453

QY      50 RGGFNGD-----PGFLGPP-----GEKGLGVPLPGYP---GRQGPKGSIGFFG 91
Db      454 NPGFQGPGRPRPGLPGEPPGLPGNGGIKGEKGNPGQGLPGLPGLKGDQGGPPGLQGNFG 513

QY      92 FPGANGKGRGTFPKPRGQRGPTGPRGERGPR---GITGKGP-----K 135
Db      514 RPLNGMKGDPLPGVPGFPMKMGSPVPGSAGEGEPGLIGPPPPGLPSPSGQSIIIK 573

QY      136 GNSGGDGPAGPPGERGNPGQPTGPPGKPPGPKDGLGHGPGQG 184
Db      574 GDAGPPGIPGQPLKGLPGQPGQLPGFTGPPGDPGRNGLPGFDGAGG 622
```

Search completed: May 3, 2004, 13:11:09  
Job time : 22.8274 secs